

SI Appendix

Supplementary Text

Details of seroarcheologic results depicted in Fig. 1B. The H1 results are from data from ref 1 (HI titres ≥ 100 against A/Hong Kong/117/77). The H2 results from data from ref 2. The H3 results come from combining data from Fig. 2 of ref 3 (HI titres of 9-15 against an equine H3N8 virus, “A/EQUI 2/Richelieu/63”, in human sera collected in 1958) with data from Fig. 1 of ref 4 (HI titres ≥ 20 against a human H3N2 virus, “A2/Hong Kong/68”, in human sera collected in 1958), with results pooled by age group and combined into a single curve.

Seroarcheology corroborates pre-1918 emergence of H1 and extinction of H3. The highest antibody titres in an age group reflect the dominant antigens of the virus responsible for their initial childhood exposure (5, 6). In sera collected in 1935, both Shope (7) and Francis and Magill (8) found that protection against H1 antigens peaked not in those born just before 1918 but in those born between 1896 and 1905 (Fig. 1 in ref 7; Chart 2 in ref 8). Rekart *et al.* (9) and Masurel and Heijntink (1) found the identical pattern forty years later: in sera collected in 1976 and 1977, respectively, HI antibody assays with H1N1 viruses revealed peak H1 protection in those born between 1896 and 1905 (9) and 1896 and 1907 (1) (Fig. 2 in ref 9; Fig. 1 in ref 1).

Specifically, Masurel and Heijntink (1) found in 728 sera collected from patients with birth years between 1883 and 1930 that the highest percentage with HI antibodies to seasonal H1N1 viruses was in those born in 1903-4 (their results with A/Hong Kong/117/77 are superimposed on the 1918 case fatality curve in Fig. 1B: “H1”). If H1 had actually been first introduced to the human population in the 1918 pandemic one would expect to find the greatest percentage of H1 antibodies in sera collected from those born in or just prior to 1918. However, ~90% those born in 1903-4 had HI titres ≥ 50 against A/Hong Kong/117/77, while just ~10% of those born closer to 1918 (birth year 1909-10) did (1).

Furthermore, there is no strong seroarcheological evidence that H3 circulated for long after ~1900 (2-4, 10-13). For example, in 913 sera collected in 1956-57, Masurel (2) found that only those born before 1897 exhibited high HI antibody titres either to human H3 or equine H3 HA antigens. And in sera collected in 1963, while 50 of 435 taken from those born prior to 1903 showed antibodies against equine H3 antigens, only 2 of 434 from those born later did (3).

There is also no evidence from mortality patterns in 1968-1969 that H3 circulated after 1900: unlike the strong protection in those born just prior to, or during, the 1889-93 pandemic, those younger than 70 (born after 1896) experienced no discernable protection from excess mortality upon exposure to the 1968 H3N2 virus (11). Attack rates in 1968-69 were also about three times higher for those born after 1899 than those born prior to 1890 (12). These observations are not consistent with continued circulation of an H3 virus up until the 1918 pandemic.

Dowdle (5) noted, based on seroarcheological findings with H1 (14) and H3 (2), that about half of those born in 1893 had been primed with the 1889 H3, and half with H1, which he assumed had emerged in 1918, 25 years after 1893. We contend that it is untenable to argue that a virus with an average annual attack rate of 20-30% in children (15) would have left 50% of 25 year-olds immunologically naïve to an H3 virus that supposedly circulated up until the 1918 pandemic: this would imply an unrealistic annual attack rate averaging just 2% over a quarter century. It seems doubtful that there could have been more than a few percent of 25-year-olds in 1918 with no prior exposure whatsoever to influenza A virus (except in the most remote locations). Dowdle's observation, on the other hand, is easily reconciled with the hypothesis that H3 was replaced by H1 around 1900, and that ~50% of those born in 1893 were not exposed to H3 antigens prior to the extinction of that virus and were primed with H1 after it emerged in ~1900.

Classical swine influenza and postpandemic H1N1. To our knowledge the earliest evidence of swine influenza in 1918 comes from Dr. Grant Munger, an inspector from the Cedar Rapids, Iowa Division of Hog Cholera Control of the Bureau of Animal Industry who observed herds of swine ill with influenza in western Illinois, in August (note 1 in ref 7).

The evidence that the classical swine virus in 1918 was a direct descendant of the pandemic virus, but the postpandemic seasonal virus was not (at least in HA), makes sense of some confusing observations from

seroarcheology—including the abrupt disappearance of antibodies to the swine virus HA in those born after about 1922 (1, 7, 8, 16, 17), even though this same cohort showed a high frequency of potent antibodies to WS/33 and other seasonal H1N1 viruses. As Andrewes *et al.* stated in 1935 (16): “The age distribution of antibodies to the porcine virus is now seen to be of some interest; the absence of neutralizing sera in children under ten may be due to complete or partial dying out of this particular strain of virus as a cause of human disease during the last decade.” Our phylogenetic results (Fig. 2) are in precise agreement with these early observations: the obvious inference is that children born after 1922 or so indeed were not exposed to the pandemic virus HA, which had by then been displaced by an antigenically and phylogenetically distinct seasonal H1N1 variant (orange star in Fig. 2), one that the seroarcheological results suggest may have been introduced near 1900.

Crucially, the precipitous decline in antibodies to H1 HA among those born from ~1918-1922, which has been put forward as a seroarcheological model for identifying pandemics (5), is observed only when antigens of classical swine influenza virus, rather than seasonal human H1N1, are used in HI assays (1, 7, 16). Our results imply that the steep drop in antibodies to the swine virus H1 actually reflects the process of a formerly widespread HA variant (the pandemic virus) rapidly going extinct as it was displaced by the reassortant seasonal H1N1 virus, with its related, but distinct, H1 HA. We believe our phylogenetic results provide a logical resolution to the sudden disappearance of antibodies to the pandemic/swine virus HA in the 1920s (7, 16) (even though H1N1 continued to circulate until 1957), as well as to the seroarcheological evidence that the HA of the seasonal H1N1 that circulated after ~1922 emerged near 1900, rather than 1918 (1).

The immunological backdrop of pandemic mortality patterns in 1918. There is seroarcheologic evidence of antibody responses against H2 antigens among the cohort born in the 1870s (2) (also see Fig. 1B). In light of our phylogenetic results, we suspect these responses likely reflect cross-reactivity to a putative 1830-1889 H1-like HA rather than to H2 per se, which our results show did not even diverge from its common ancestor with H5 until ~1875 (Figs. S13 and S14). Masurel (14) showed that >70% of those born between 1867-1886 had HI antibody titres ≥ 100 to H1 HA antigens of A/Swine/15/1930 (Fig. 1 in ref 14). This is a higher percentage than in the slightly younger age group exposed in childhood to the 1889 virus (14), suggesting the pre-1889 age groups were indeed exposed to H1-like antigens prior to the emergence of the 1889 virus (see Fig. 1B). Shope (7) found the same pattern in the 1930s: sera from 100% of those born before 1875 neutralized A/Swine/15/1930 (note the bimodal distribution in Fig. 1 of ref 7, with peak percentages of protection against H1 in those born in 1896-1905 and those born prior to 1875, i.e. either side of the 1889 pandemic).

There was also much higher mortality among >50-year-olds in 1889-90 than there was in 1918, strongly suggesting that those born before ~1840 experienced childhood exposure to a non-H3, non-N8 virus. Altogether, it seems highly plausible that an H1N1-like virus was associated with both 1830 and 1918, and that an H1-like HA circulated from 1830-1889, leaving only a short window from ~1889-1900 when an H1-like virus did not circulate in humans (Fig. S13). Much the same thing happened after 1918, with H1N1 circulating until 1957, re-emerging in 1977 and 2009, and H1N1 being absent only between 1958-1976. As shown in Fig. 1, the disappearance of an H1 HA left the ~1889 cohort with a sharp peak in heterosubtypic (with respect to H1N1) anti-HA and NA antibodies (H3 and N8, respectively), while slightly older and younger groups exhibited marked H1 or H2 HA protection.

We speculate that the 1830 H1N1-like virus acquired N8 by reassortment in or shortly before 1847, leading to that pandemic, which coincided with an extensive epizootic in horses all across Europe (18). This putative H1N8 variant may have remained as the seasonal strain until the 1889 pandemic, which would account for the evidence of antibodies to H1/H2 in those born between 1847 and 1889. Those primed by such a virus, with a homosubtypic HA and heterosubtypic NA, would be expected to have suffered intermediate excess mortality in 1918 (higher than the cohort primed by the putative 1830 H1N1-like virus, but lower than the cohort whose initial exposure was to the 1889 H3N8 virus).

The 1889-93 pandemic, the first in modern times, had a global scope not seen in prior influenza pandemics, and it was second only to 1918 in severity over the last two centuries. Revolutionary developments in transportation, especially modern steamships that led to an enormous increase in the volume and speed of intercontinental maritime travel, and railway networks that connected port cities and inland populations, allowed influenza to move between and across continents at unprecedented speeds (19). In light of this, as well as the high

incidence of disease measured during several successive waves (19), it is likely that most individuals born shortly before and during 1889-93 were primed by the 1889 pandemic virus.

Fan *et al.* (20) showed in a mouse model that M2 immunity reduces viral replication in the lungs during the entire course of infection. If so, then absence of natural immunity to M2 would likely increase viral replication in the lungs. This may be a contributing factor for the general severity of the 1918 pandemic. Our phylogenetic analyses make it clear that a new M1/2-encoding segment was acquired just prior to the pandemic from an avian virus. The prior M2 may have been so divergent that cross-protection with the 1918 M2 was poor, leading to a more lethal virus until herd immunity to the new M2 was achieved. Recovery of pre-1915 virus sequence could permit a test of this hypothesis, but the protection of the oldest age groups in 1918 suggests to us that lack of immunity to M2 or other minor antigens (like NP) may not, on its own, have been a decisive factor in 1918.

Cohort-specific immunity in 1918 and other pandemic years. The idea that young adults suffered unusually severe outcomes in 1918-1920 because of childhood exposure to H3 and N8 antigens fits with the observation that they died primarily of typical post-influenza complications. Any individual with a poor immune response to a current influenza infection is at higher risk of severe or fatal disease. Indeed, the logic behind this argument is no different than that underpinning the enterprise of protecting individuals from severe and fatal influenza-related disease by eliciting protective immune responses with vaccines. Since prior exposure to H3 protected the 1889-1900 cohort against H3N2 in 1968-70 (11), it stands to reason that the same age group would have been especially vulnerable to the antigenically mismatched virus in 1918, compared to other age groups exposed in childhood to N1 and/or H1 antigens. This is similar in kind if not degree to individuals <70 in 1968, <20 in 1977, and <52 in 2009 (11, 15). These younger age groups, primed in childhood with heterosubtypic HA or NA antigens, all suffered more excess mortality than older age groups who were apparently protected by childhood exposure to viruses with partial or complete overlap in HA and NA subtypes.

Were clinical attack rates in 1918 of about 28% overall and 30% in adults 20-40 years of age (21) consistent with the possibility of a pre-pandemic interval of H1 HA circulation? This is lower than observed in some interpandemic (seasonal) epidemic years when virtually all adults have had prior exposure. Prior exposure to H1 HA would only be expected to provide sterilizing immunity (and thus produce very low attack rates) against homologous HA variants, and even a year or two of antigenic drift is often sufficient to create heterologous HA. (That is why vaccines against seasonal influenza A viruses are usually reformulated every year or two). Our phylogenetic results suggest that the circulating pre-pandemic H1 HA antigens would have been as much as 22 years diverged from the pandemic HA (11 years each of independent evolution from the MRCA at 1907 to the pandemic and seasonal lineages evidently circulating in 1918). The 1918 HA would have been 11 years diverged from the putative H1 HA infecting individuals back in 1907. Only a very small proportion of the population would be expected to have been exposed to an HA homologous to the pandemic HA. What was unusual about the 1918 pandemic was not its attack rate (which was typical and consistent with a pre-pandemic interval of H1 HA transmission) but its severity, particularly in young adults.

Possible limitations of the clock model. As mentioned in the main text, it is possible that hidden or reverted host states within the data sets we analyzed may affect our conclusions about the direction of host jumps and the timing of these events. For one thing, as in ref 24, we excluded from our analyses mammalian clades nested within other mammalian clades, which cannot be modeled with the current implementation of the HSLC method. The only such case in the present analysis is the 2009 pH1N1 virus, which originated in swine and moved to humans in or shortly before 2009. We can see no way in which this exclusion would affect our estimates of the local clock rate in the human H1N1 lineage depicted in Figs. 2, S1, and so on: events occurring after 2009 could not have affected inferences based on sequences from earlier years. Moreover, we do not believe there is any chance that the distinct swine and human clades apparent in our trees, from the 1930s onwards, contain hidden host states. There is ample epidemiological evidence that these IAV lineages circulated continuously and separately in each host (with the exception of occasional ‘dead-end’ jumps that are easy to visualize on the trees and exclude).

More problematic are the longer interior branches in the deeper parts of the tree, prior to the 1930s samples in each lineage. Again, however, there is strong epidemiological evidence that the classical swine influenza virus

circulated in each year from 1918 until the first isolations of the virus in the 1930s, so there seems little doubt that that branch is correctly specified in our analyses. Similarly, most of the stem branch leading up to the seasonal human H1N1 clade is almost certainly correctly specified as human since H1N1 circulated from at least 1918 until the first human isolates of the 1930s.

Least clear is the specification of the branches between 1918 and the common ancestor of the pandemic and seasonal H1 lineages in ~1907. Nevertheless, given the topology of the tree, the most parsimonious scenario is clearly that the human pandemic virus and the seasonal H1N1 lineage shared a common ancestor in humans (Fig. S12A). Other scenarios require one or more additional host jumps (Fig. S12B-F); they also are less consistent with (i) the seroarcheologic evidence entry of H1 antigens into the human populations between about 1896 and 1907 (1-4, 7-9, 10-12, 15); (ii) the evidence of several years of mammalian transmission indicated by the high uracil content of the 1918 HA gene; (iii) the epidemiological evidence that influenza was new to swine in 1918 (7, 25); and (iv) the lack of H3 HA antibodies and lack of protection during the 1968 H3N2 pandemic in those born between ~1900 and 1918, suggesting that a non-H3 IAV subtype circulated in humans in that period (1-4, 10-13). These independent lines of evidence suggest (but, we recognize, do not definitively prove) that an H1 HA circulated in humans a decade or so prior to 1918. Moreover, even if one of these more complex scenarios indeed obtained, the timing of the ancestor of the pandemic and seasonal H1 lineages might shift by a few years but the overall pattern of a deep divergence between them would remain, as would the close relationship between the 1918 HA and the swine H1N1 HA genes (much closer than the pandemic and seasonal HA genes are to each other).

It is also worth noting that while the 1930s and later swine H1N1 lineage and the 1930s and later human H1N1 lineage were constrained to be monophyletic in our analyses, both of these assumptions were carefully validated with non-clock trees. Furthermore, the 1918 human sequences were not enforced to be monophyletic with other human strains. Thus, the monophyly constraints imposed are highly unlikely to have biased the results in any substantive way.

Antigenic imprinting and seasonal influenza mortality patterns. The 1968 H3N2 pandemic was notable for its lack of severity in the elderly (11, 12). Now that the elderly cohort from 1968—whose childhood exposure to H3 in 1889-1900 provided protection to H3N2 (12, 13)—has been supplanted by an elderly cohort primed with H1N1, the pattern has reversed: the same virus is now noteworthy for its severity in the elderly and its relative mildness in younger patients. The current pattern of high H3N2-caused mortality in older adults (22, 23) is difficult to reconcile with faltering immunity in the aged: similarly aged adults, evidently primed as children by an H3 HA, had good protection in 1968-69, with no excess mortality. The pattern also seems at odds with the idea that properties intrinsic to this strain of IAV make it particularly dangerous for elderly patients: H3N2 was not inherently virulent in the elderly when it first emerged. While it is possible that intrinsic properties of the virus or weakened immunity in the elderly somehow account for this reversal, cohort immunity shaped by childhood exposure to homosubtypic or heterosubtypic antigens seems more likely in light of our findings. (For H1N1, conversely, the current elderly cohort, who were exposed to H1N1 as children, appear to be protected: we surmise that their childhood exposure to H1N1 reduces the effective number of susceptibles and reduces the frequency of severe postinfluenza complications leading to death.)

Influenza-related hospitalizations are about twice as high in the U.S. during H3N2-dominated seasons (22), and H3N2-related all-cause mortalities are more than 10 times higher than for H1N1 (23). Deaths among the elderly are disproportionately numerous with H3N2 compared to H1N1. During the 1990s, the estimated annual number of H1N1-related deaths (all cause) among 1-4 year-olds was 34; for H3N2 it was 103 (23). For each H1N1 death in the 1-4 age group, there were 57 deaths in those ≥ 65 . However, for H3N2 this ratio was 1:338, almost six times greater. (The numbers of annual H1N1-related and H3N2-related deaths in those ≥ 65 were 1,954 and 34,866, respectively (23), or ~18 H3N2 deaths per H1N1 death in the elderly group.)

This suggests to us a greatly increased risk of death when elderly patients are infected by a seasonal virus with HA and NA glycoproteins heterosubtypic to those of their first exposure in childhood. Conversely, the 5-49 year-old group (who, because of the emergence of an H3N2 virus in 1968 are expected to have relatively good protection against H3N2 compared to those >65 , and relatively poor H1N1 protection) accounted for 17.7% of H1N1 related deaths but just 4.2% of H3N2-related deaths. Therefore, as in 1918 and other pandemics, childhood immunity appears to have a strong effect in shaping seasonal influenza mortality patterns. Thompson *et al.* (23)

describe H3N2 as the “most virulent of the recently circulating influenza viruses”. We hypothesize that this ‘virulence’ is due largely (or perhaps even completely) to host immunity, not to intrinsic properties of the H3N2 virus.

Antigenic imprinting and H5N1 versus H7N9 mortality patterns. We compared patterns of mortality due to avian-origin H5N1 and H7N9 in different birth year cohorts using H5N1 fatality data from Indonesia in 2005-2005 (26) and H7N9 fatality data from China in 2013 (27). Although others have noted that H5N1 tends cause severe disease and death primarily in younger age groups and H7N9 in older age groups (27), to our knowledge birth year (rather than age at time of infection) in its connection to group 1 versus group 2 HA exposure has not been considered previously.

As depicted in Fig. S15A, 100% of H5N1 fatalities in 2005-2006 in Indonesia occurred among patients born in 1968 or later, after the emergence of H3N2 that occasioned a switch from group 1 HA to group 2 HA. In contrast, 85% of H7N9 fatalities in China in 2013 occurred among those born prior to 1968. We performed a 2x2 Chi-square test for each subtype as follows: we compared the observed number of fatalities in those born prior to 1968 (exposed in childhood to a group 1 HA of either the H1 or H2 subtype) and those born in 1968 and later (exposed in childhood primarily to a group 2 HA of the H3 subtype). We then calculated the expected number of fatalities in each of these two cohorts if cases were evenly distributed, correcting for the different population size of each cohort by using census data for Indonesia in 2006 or China in 2013 (28).

For H5N1, the observed and expected post-1967 fatalities were 41 and 28, respectively, and the observed and expected pre-1968 fatalities were 0 and 13, respectively (Chi-square=15.4, df=1, two-tailed P<0.0001). For H7N9, the observed and expected post-1967 fatalities were 5 and 23, respectively, and the observed and expected pre-1968 fatalities were 29 and 11, respectively (Chi-square=19.7, df=1, two-tailed P<0.0001). Note that the age pattern is reversed for the different viruses, consistent with initial childhood exposure to a ‘group-mismatched’ HA greatly increasing mortality risk, and heterosubtypic but ‘group-matched’ HA providing near complete protection from fatal disease upon exposure to a novel, avian-origin influenza A virus. Combining both H5N1 and H7N9 fatalities suggests that ~93% of fatalities have occurred in individuals exposed to an avian virus with an HA mismatched to their childhood HA group.

This can be seen in Fig. S15B, in which census data (28) are used to account for the proportion of the total population (in each country) contained in each 10-year birth year bin. (Values above 0 indicate a greater number of fatalities than if the fatalities were distributed in proportion to the underlying demographic distribution.) The distributions for the two viruses are near-mirror images, centered on the switch from group 1 HA to group 2 HA in 1968. Moreover, much of the ~15% mortality in the right tail of the H7N9 distribution (in birth years of 1968 or later) is plausibly due to the re-emergence of H1 in 1977, which we hypothesize may make a minority of those born after 1977 (in particular those born close to the so-called ‘pandemic’ in 1977) more susceptible to severe disease when infected with a group-mismatched (H7) HA virus. If this is correct, then the co-circulation of H3N2 and H1N1 since 1977 may shape severity outcomes in those born from ~1977 onward such that those initially exposed to the group 1 virus (H1N1) are more susceptible later to H3N2 and H7N9 and those initially exposed to the group 2 virus (H3N2) are more susceptible later to H1N1 and H5N1. It may be possible to test this hypothesis with epidemiologic or immunologic approaches.

Supplementary References

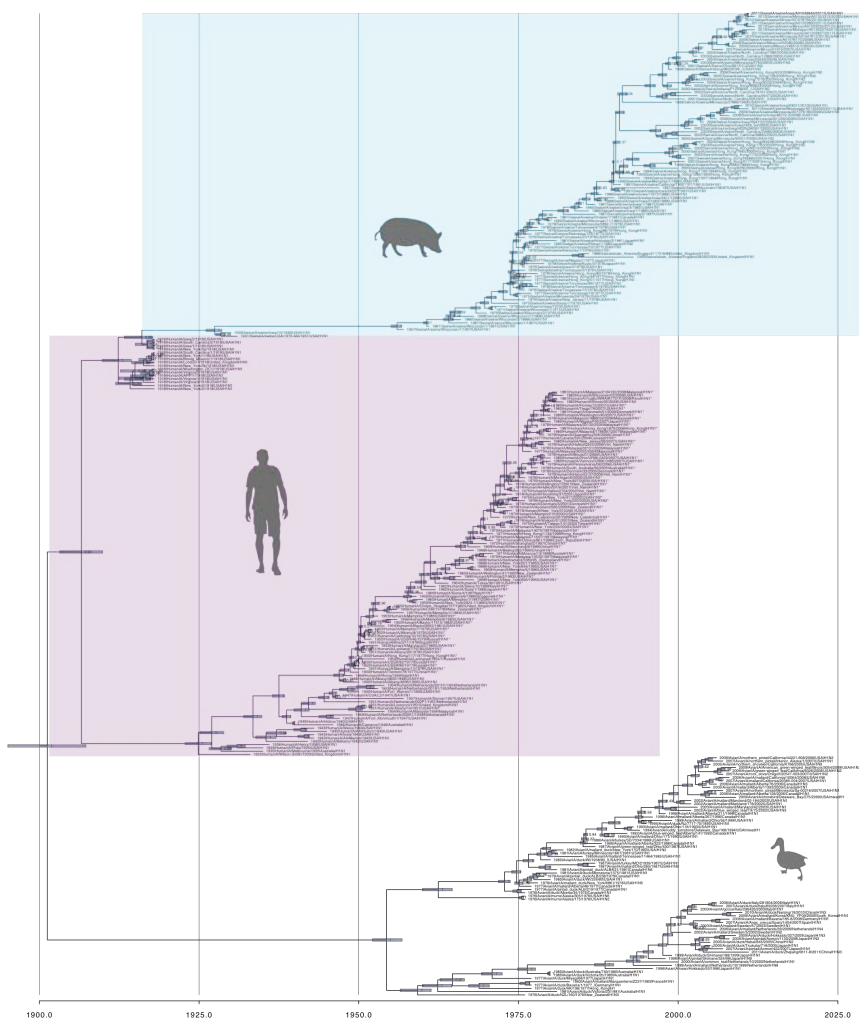
1. Masurel N, Heijntink RA (1983) Recycling of H1N1 influenza A virus in man—a haemagglutinin antibody study. *J Hyg (Lond)* 90(3):397-402.
2. Masurel N (1969) Relation between Hong Kong virus and former human A2 isolates and the A-EQU12 virus in human sera collected before 1957. *Lancet* 1(7601):907-910.
3. Masurel N, Mulder J (1966) Studies on the content of antibodies for equine influenza viruses in human sera. *Bull World Health Organ* 34(6):885-893.
4. Davenport FM, Minuse E, Hennessy AV, Francis T (1969) Interpretations of influenza antibody patterns of man. *Bull World Health Organ* 41(3):453-460.

5. Dowdle WR (1999) Influenza A virus recycling revisited. *Bull World Health Organ* 77(10):820-828.
6. Francis T (1960) On the doctrine of original antigenic sin. *Proc Am Philos Soc* 104(6):572-578.
7. Shope RE (1936) The incidence of neutralizing antibodies for swine influenza virus in the sera of human beings of different ages. *J Exp Med* 63(5):669-684.
8. Francis T, Magill TP (1936) The incidence of neutralizing antibodies for human influenza virus in the serum of human individuals of different ages. *J Exp Med* 63(5):655-668.
9. Rekart M, Rupnik K, Cesario TC, Tilles JG (1982) Prevalence of hemagglutination inhibition antibody to current strains of the H3N2 and H1N1 subtypes of influenza A virus in sera collected from the elderly in 1976. *Am J Epidemiol* 115(4):587-597.
10. Kendal AP, Minuse E, Maassab HF, Hennessy AV, Davenport FM (1973) Influenza neuraminidase antibody patterns of man. *Am J Epidemiol* 98(2):96-103.
11. Simonsen L, Reichert TA, Miller C (2004) The virtues of antigenic sin: consequences of pandemic recycling on influenza-associated mortality. *International Congress Series* 1263:791-794.
12. Schoenbaum SC, Coleman MT, Dowdle WR, Mostow SR (1976) Epidemiology of influenza in the elderly: evidence of virus recycling. *Am J Epidemiol* 103(2):166-173.
13. Davenport FM, Minuse E, Hennessy AV, Francis T (1969) Interpretations of influenza antibody patterns of man. *Bull World Health Organ* 41(3):453-460.
14. Masurel N (1976) Swine influenza virus and the recycling of influenza-A viruses in man. *Lancet* 2(7979):244-247.
15. Bodewes R, et al. (2011) Prevalence of antibodies against seasonal influenza A and B viruses in children in Netherlands. *Clin Vaccine Immunol* 18(3):469-476.
16. Andrewes CH, Laidlaw PP, Smith W (1935) Influenza: Observations on the recovery of virus from man and on the antibody content of human sera. *Br J Exp Pathol* 16(6):566-582.
17. Davenport FM, Hennessy AV, Drescher J, Mulder J, Francis T (1964) Further observations on the relevance of serologic recapitulations of human infection with influenza viruses. *J Exp Med* 120(6):1087-1097.
18. Law J (1874) *Report of the Commissioner of Agriculture, 1872: Influenza in Horses* (Government Printing Office, Washington).
19. Patterson KD (1986) *Pandemic influenza, 1700-1900: A study in historical epidemiology*. (Rowan & Littlefield, Totowa, N. J.).
20. Fan J, et al. (2004) Preclinical study of influenza virus A M2 peptide conjugate vaccines in mice, ferrets, and rhesus monkeys. *Vaccine* 22 (23-24):2993-3003.
21. Frost WH (1919) The epidemiology of influenza, 1919. *Public Health Rep* 34 (33):1823-1861.
22. Thompson WW, et al. (2004) Influenza-associated hospitalizations in the United States. *JAMA* 292(11):1333-1340.
23. Thompson WW, et al. (2003) Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* 289(2):179-186.
24. Worobey M, Han GZ, Rambaut A (2014) A synchronized sweep of the internal genes of modern avian influenza virus. *Nature* doi:10.1038/nature13016.
25. Koen JS (1919) A practical method for the field diagnoses of swine diseases. *Am J Vet Med* 14(9):468-470.
26. Sedyaningsih ER, et al. (2007) Epidemiology of cases of H5N1 virus infection in Indonesia, July 2005-June 2006. *J Inf Dis* 196(4):522-527.
27. Dudley JP, MacKay IM (2013) Age-specific and sex-specific morbidity and mortality from avian influenza A(H7N9). *J Clin Virol* 58(3):568- 570.
28. <http://www.census.gov/population/international/data/idb/informationGateway.php>

Supplementary Figures S1 to S15

Fig. S1. H1 MCC tree and circulating diversity in 1918. **(A)** H1 MCC tree (the full version of the tree depicted in Fig. 2). Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk. **(B)** Years of genetic diversity circulating in *HA* in 1918 (i.e. the time to the most recent common ancestor of the pandemic and seasonal *HA* lineages sampled at each time point) compared to circulating diversity in other years (both for seasonal H1N1 and for later pandemics). Circles indicate median node dates and gray rectangles indicate 95% CIs.

A



B

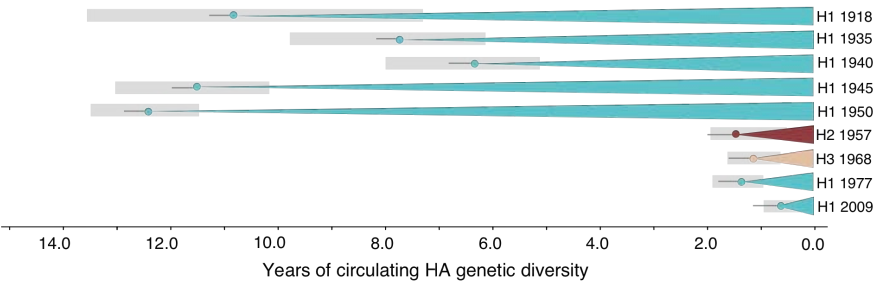


Fig. S2. H1 MCC tree (1930s laboratory strains and 1918 short sequence fragments removed). Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk.

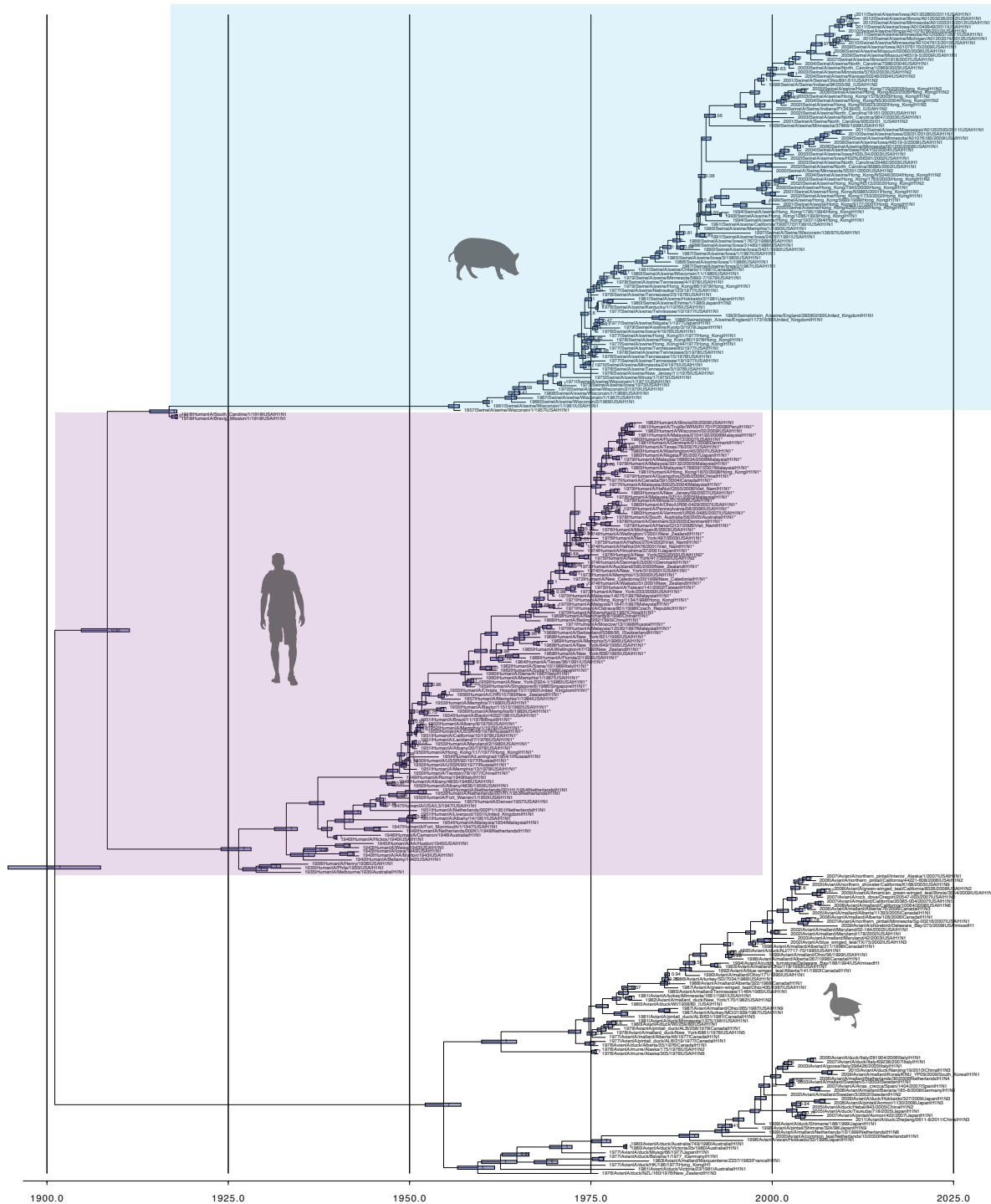


Fig. S3. H1 MCC tree (1930s laboratory strains and 1918 short sequence fragments removed, maximum of one sequence/year/host, post-1957 human H1N1 sequences removed). Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown.

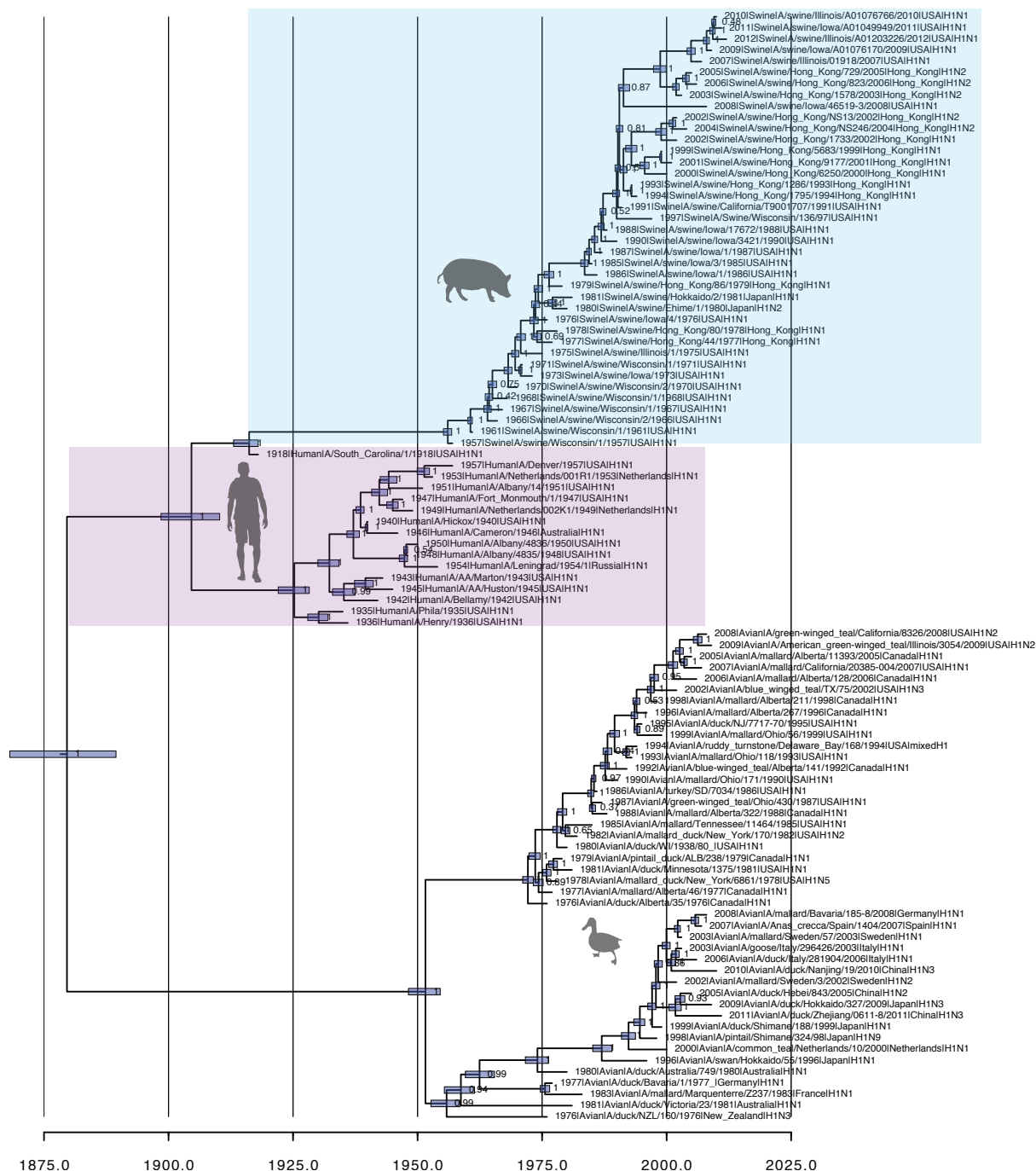


Fig. S4. H1 MCC tree (HA2/stalk domain only). Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk.

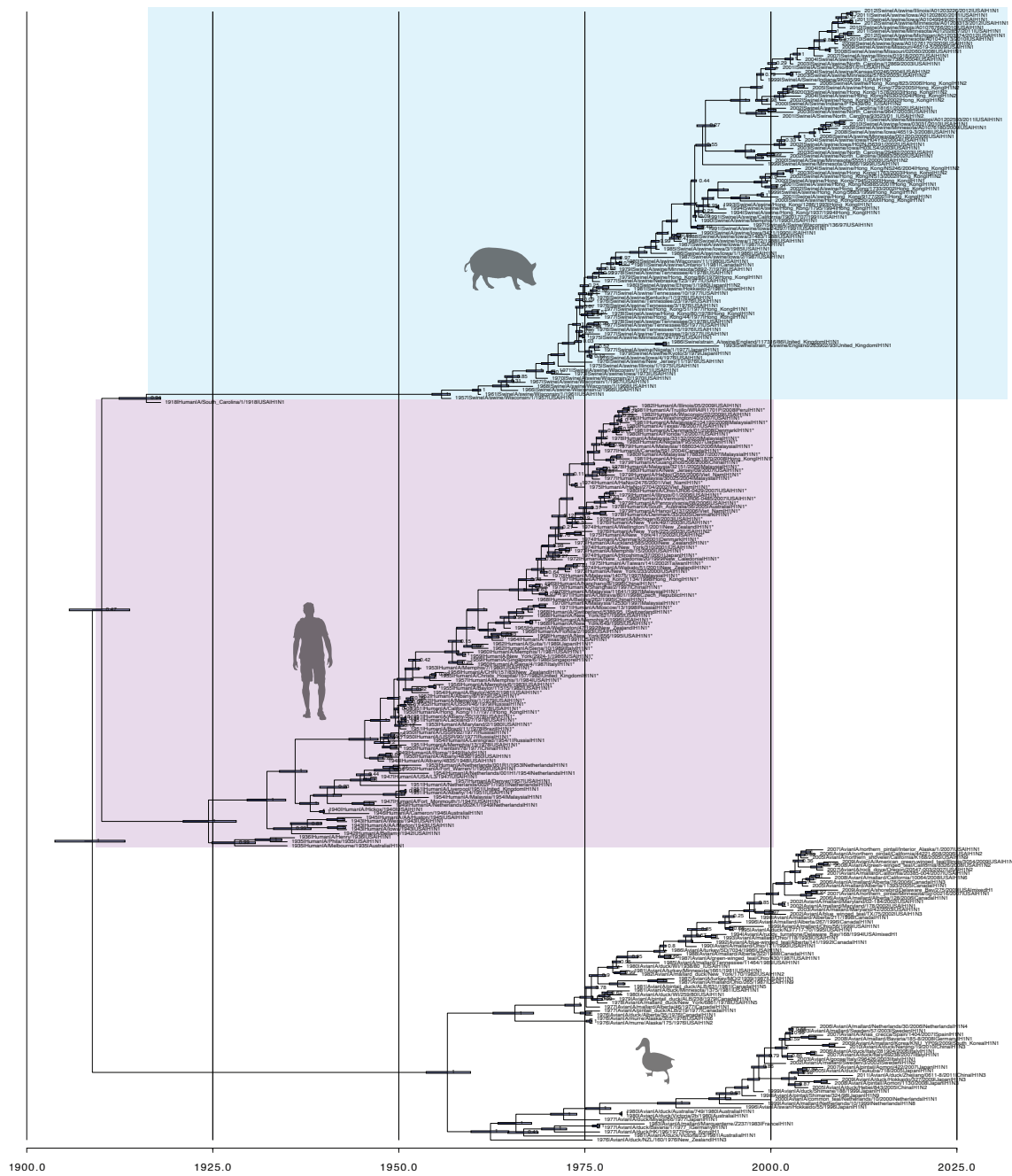


Fig. S5. Branch-site REL analyses to test for episodic diversifying selection. The branches are colored to depict the proportion of substitutions along each branch that are under purifying selection (with $dN/dS < 1$: blue), the proportion evolving neutrally (with $dN/dS = 1$: gray), and the proportion under diversifying selection (with $dN/dS > 1$: red). Almost every site in every branch evidently evolved under purifying selection, and none of the branches exhibited significant evidence of episodic diversifying selection.

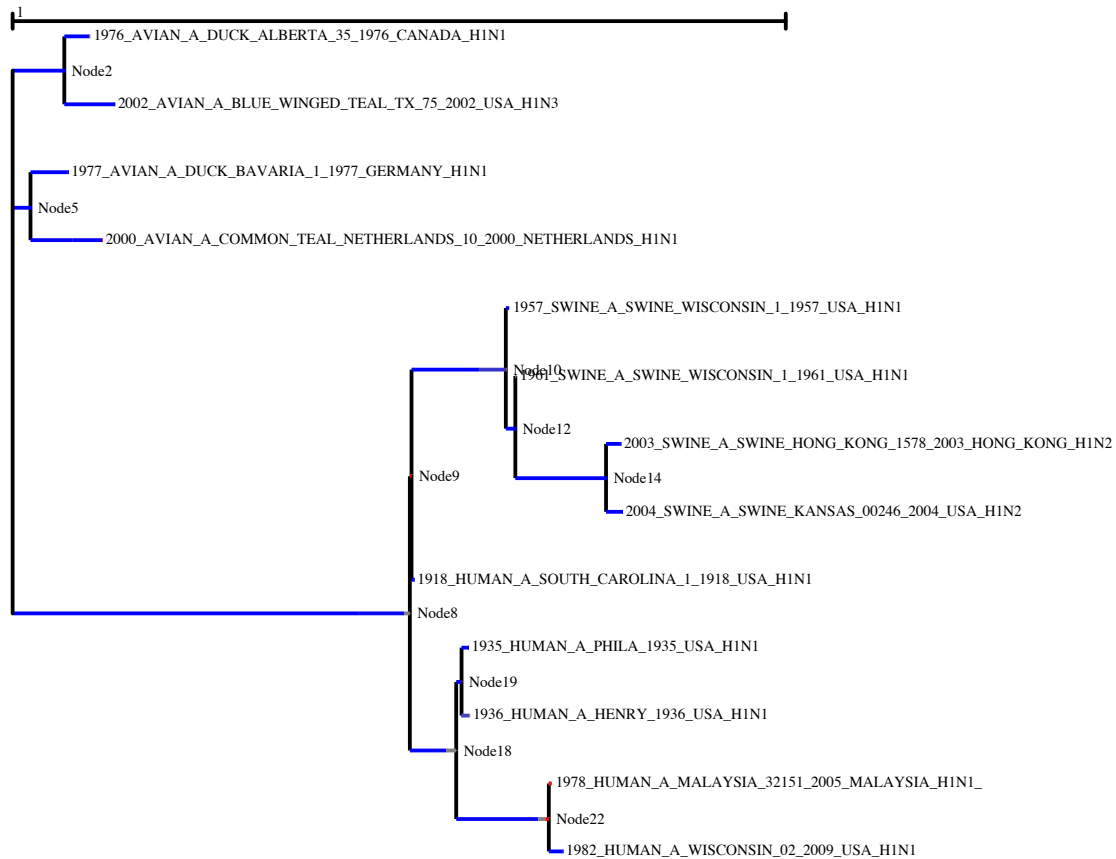
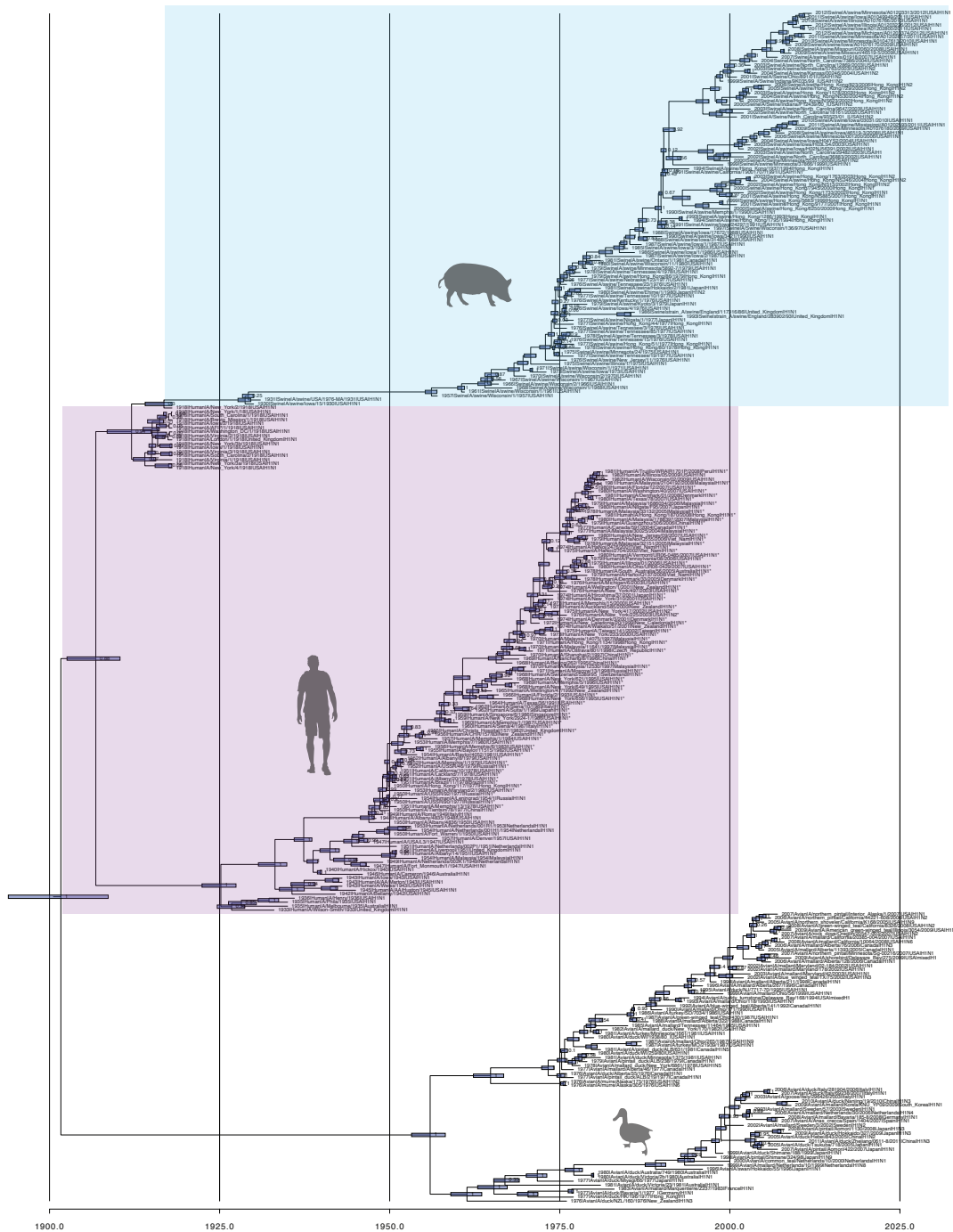
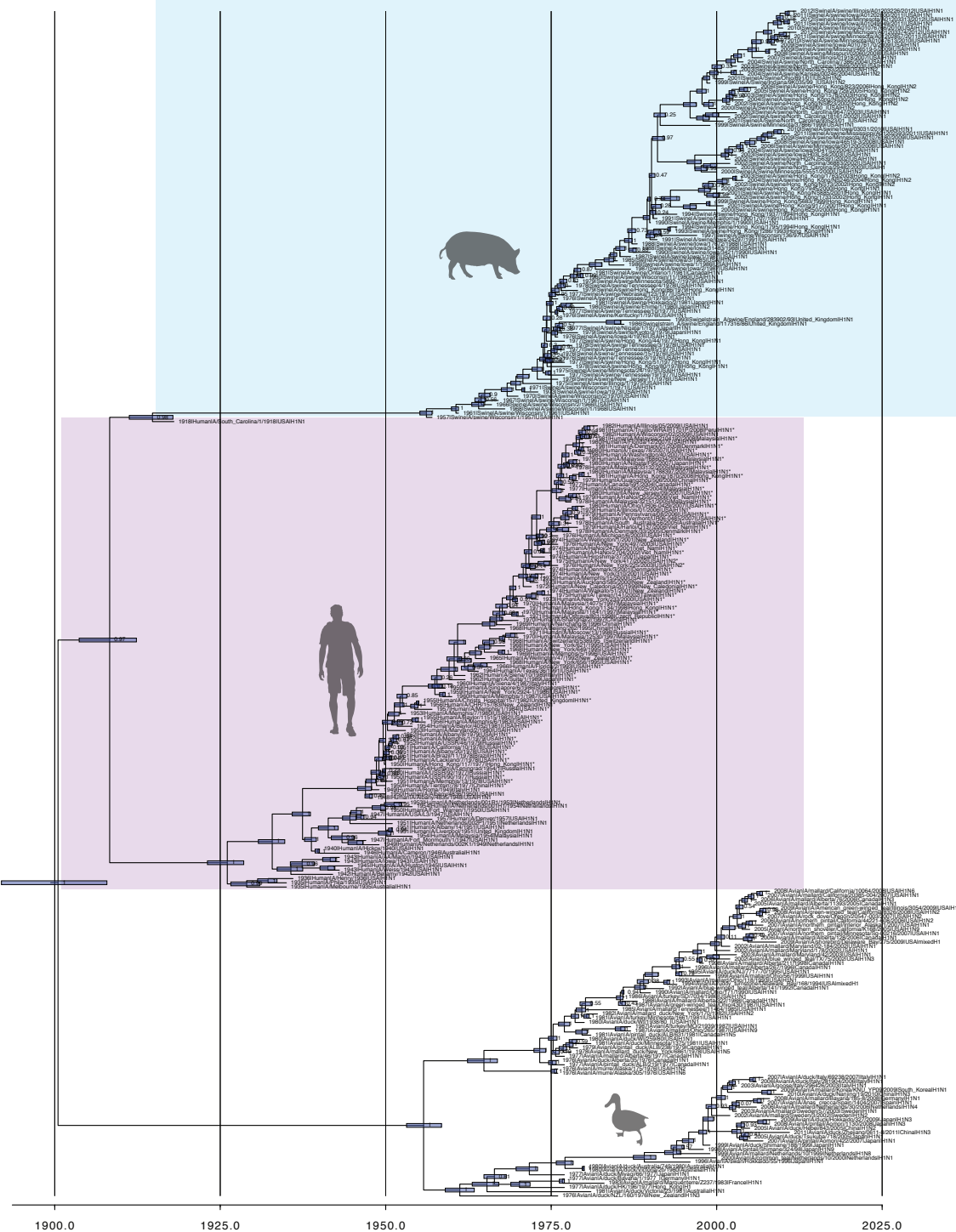


Fig. S6. H1 MCC tree (A) 1st and 2nd codon position sites removed (i.e., 3rd sites only). (B) 3rd sites only, plus 1930s laboratory strains and 1918 short sequence fragments removed. (C) Same as B but including only 3rd position sites at which substitutions were synonymous (silent sites). In all cases branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk.

A



B



C

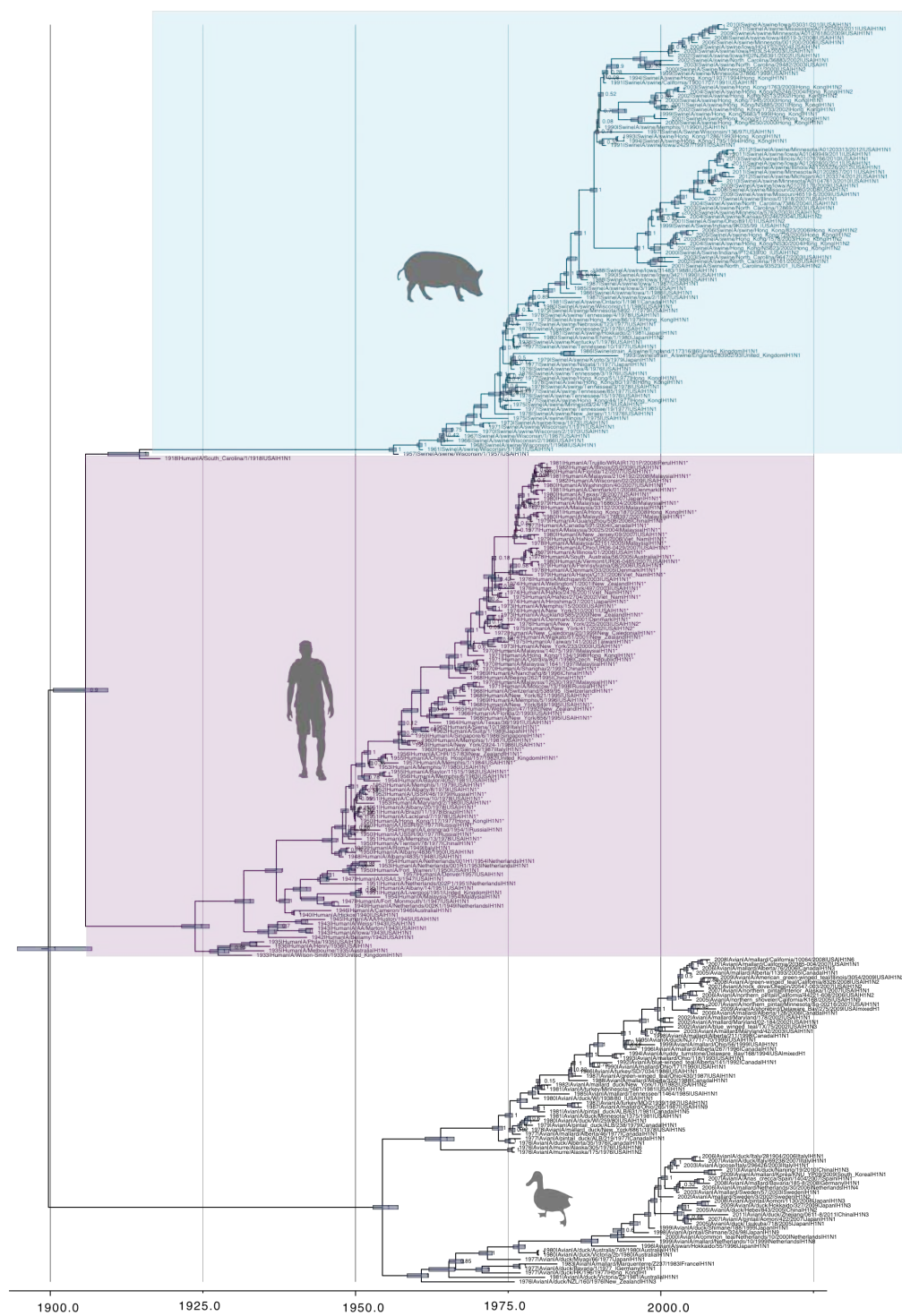


Fig. S7. Uracil (U) content patterns. The 95% CI of avian U content is shown for each segment with a gray rectangle. U content versus year of sampling is shown by magenta diamonds for the 1918 human H1N1 sequences of (A) N1 *NA*, and (B) H1 *HA*. The *P* value in panel B reflects a test of whether the U content of the 1918 *HA* is significantly higher than the avian range based on the rate of change observed in the H3N2 lineage.

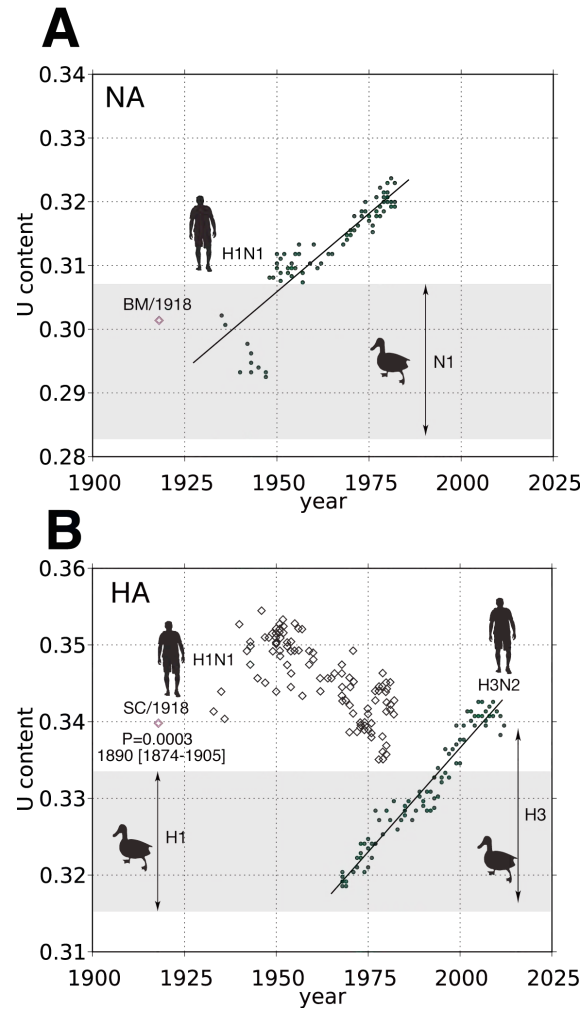


Fig. S8. N1 MCC tree. Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1N1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk. The 95% CI for the within-human H1 diversity is shown for comparison.

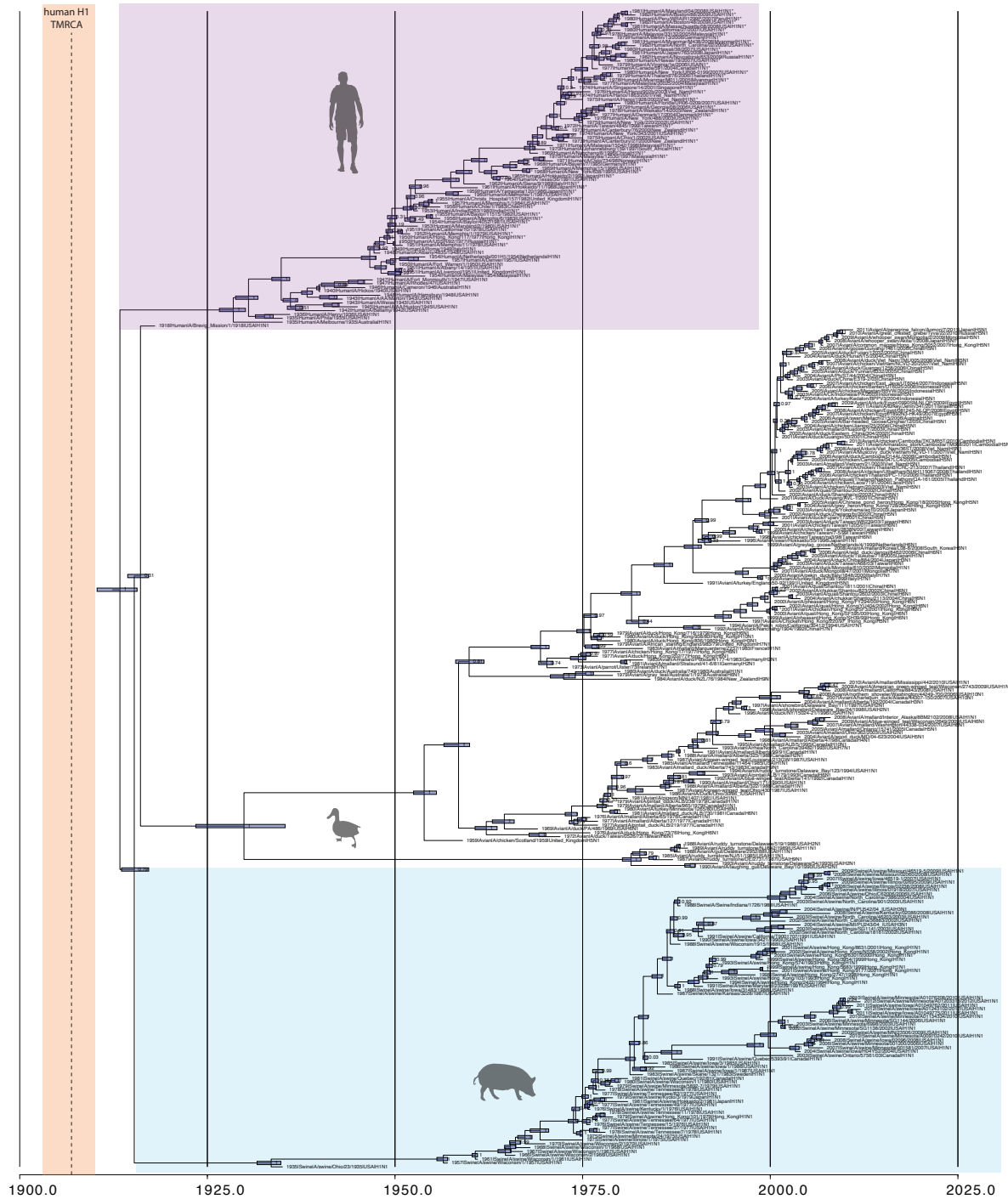


Fig. S9. *PBI* MCC tree. Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1N1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk. The 95% CI for the within-human H1 diversity is shown for comparison.

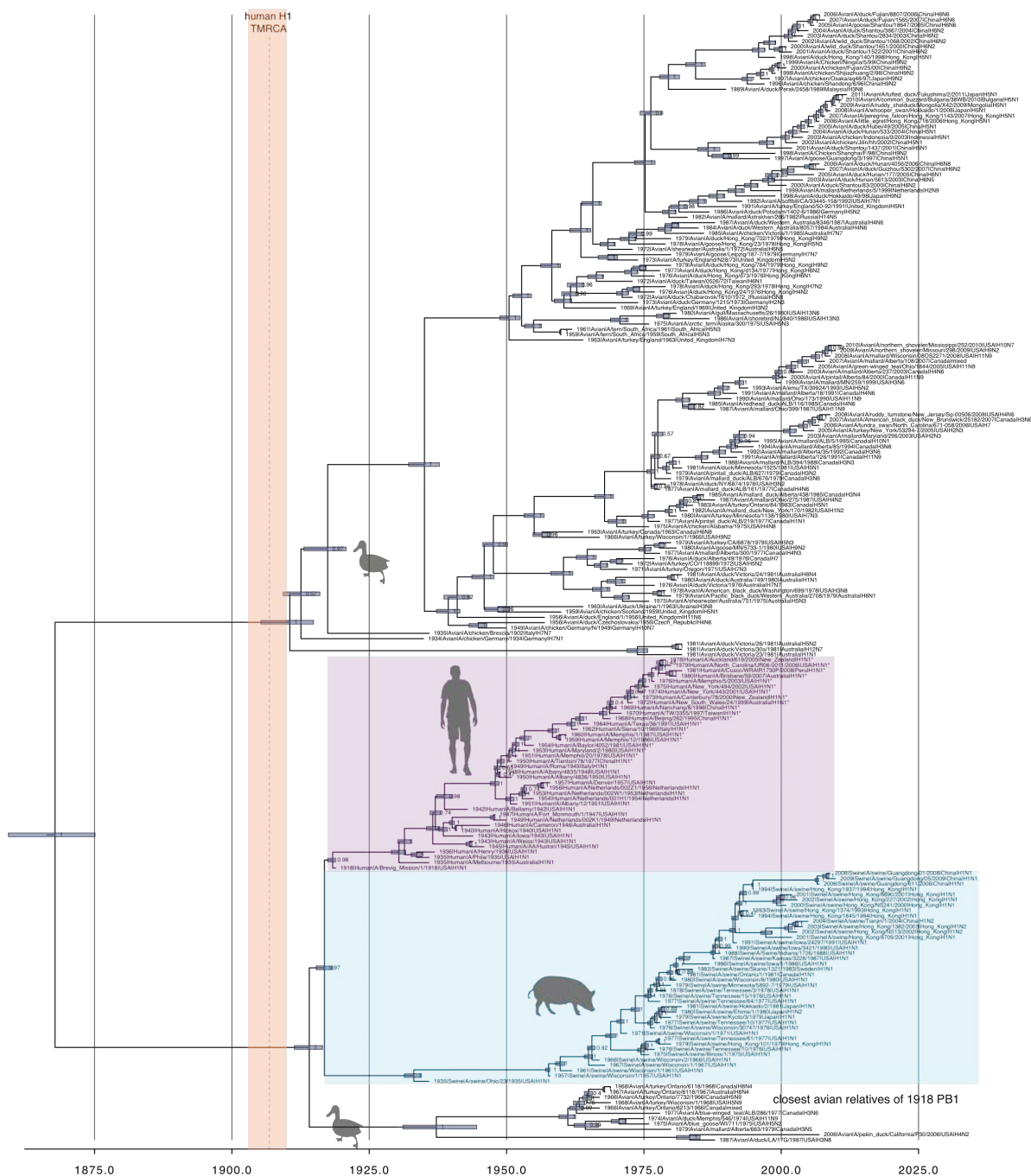
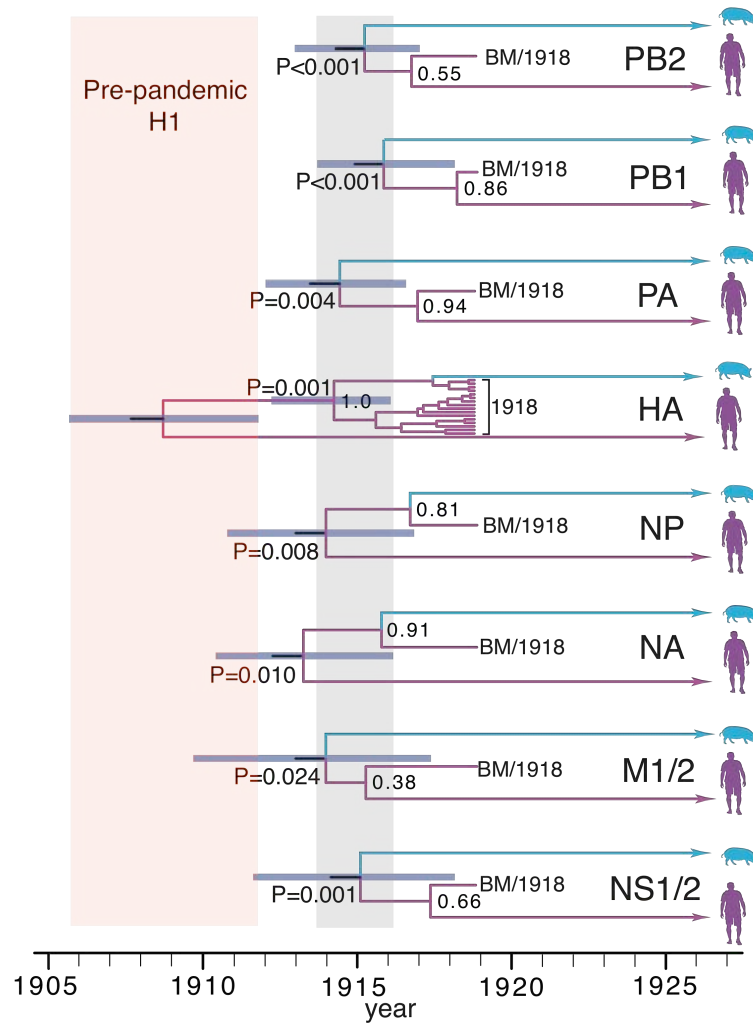
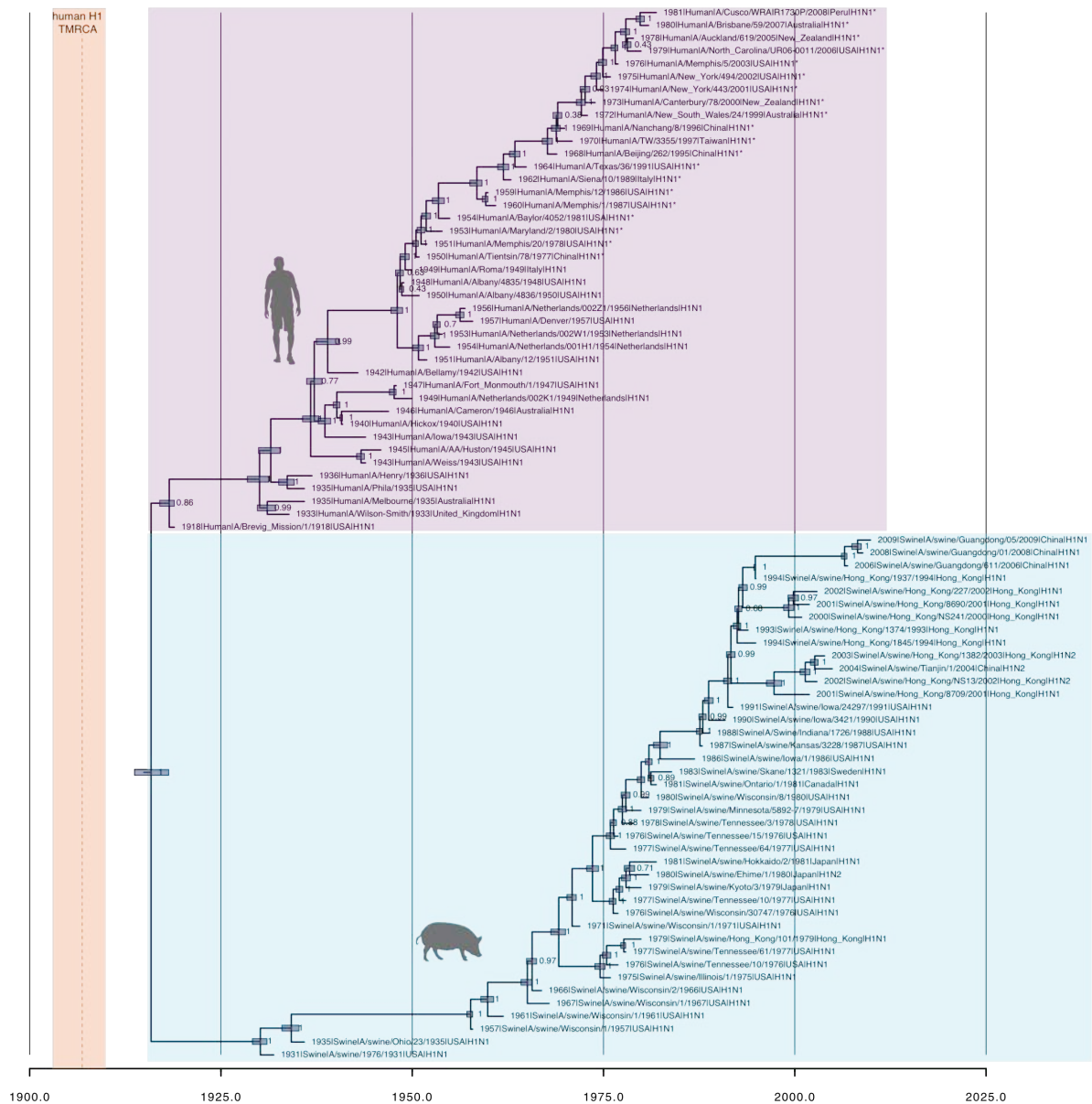
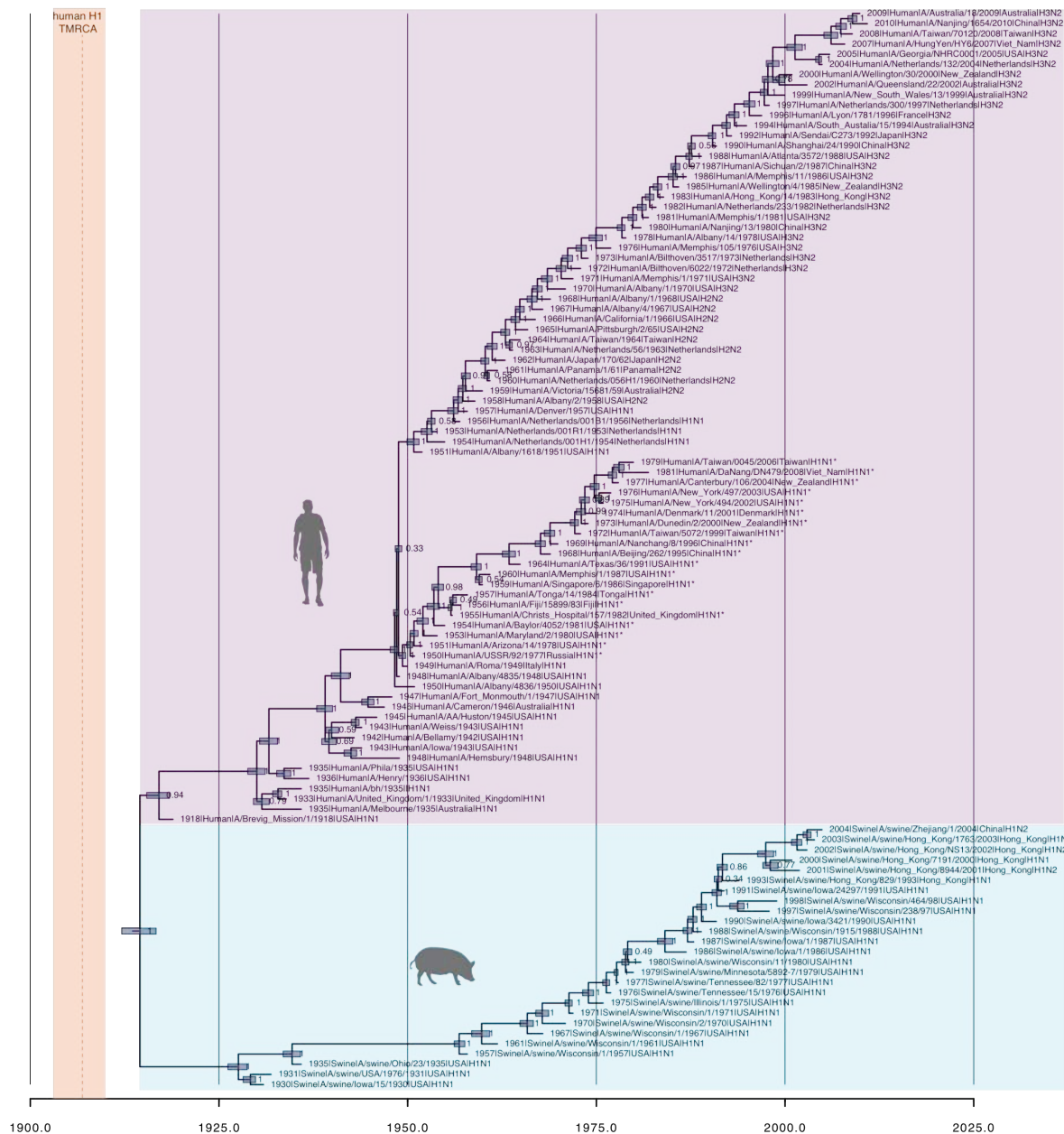


Fig. S10. H1 in humans predates all other genes in the 1918 pandemic virus. The gray bar indicates the window of overlap among the 95% CIs of the TMRCA of human and swine H1N1 for the different segments. The TMRCA of human H1 significantly predates that of each of the other genes. The TMRCA of the human pandemic and season lineages also predates the TMRCA of the classical swine/pandemic human H1 node.

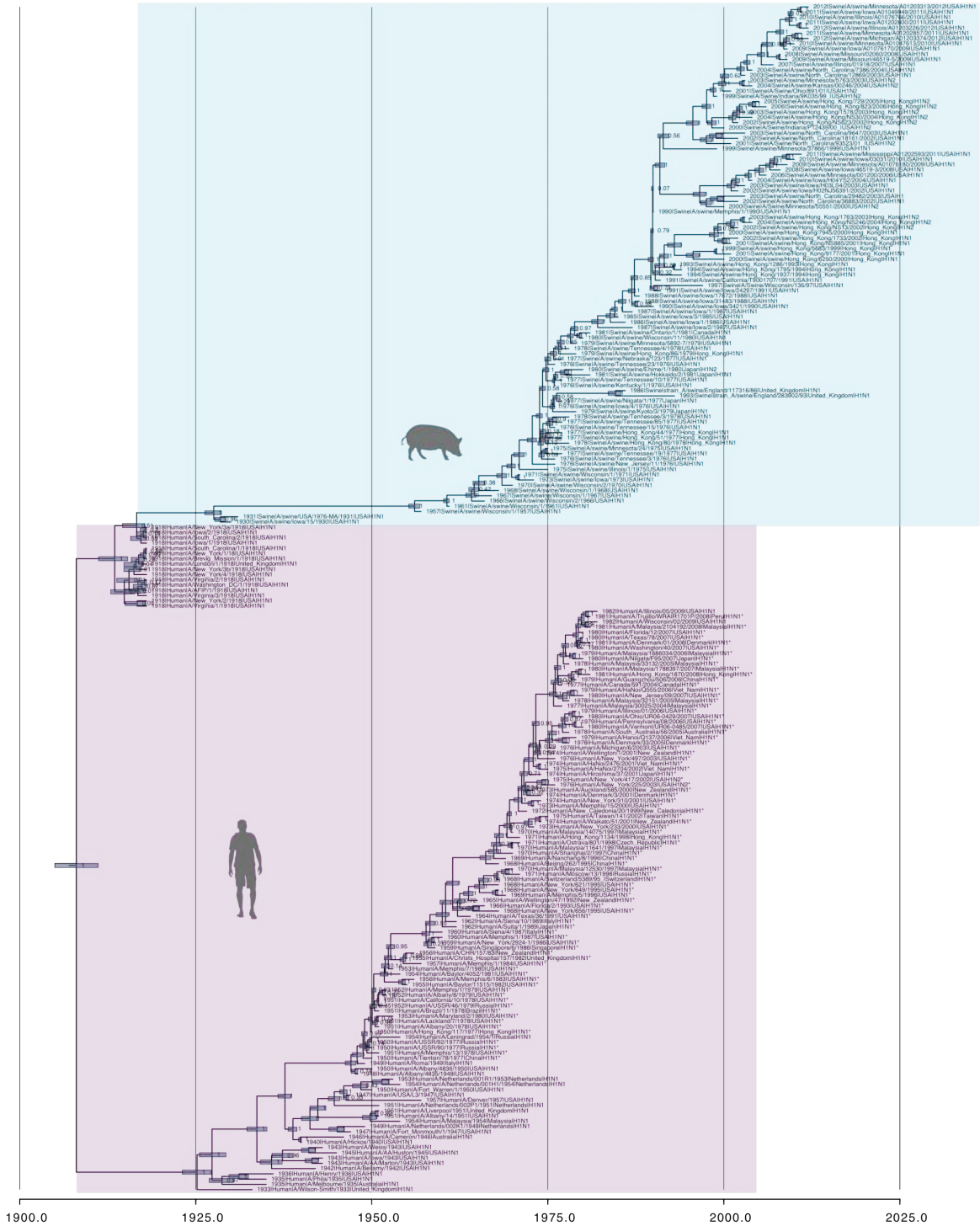


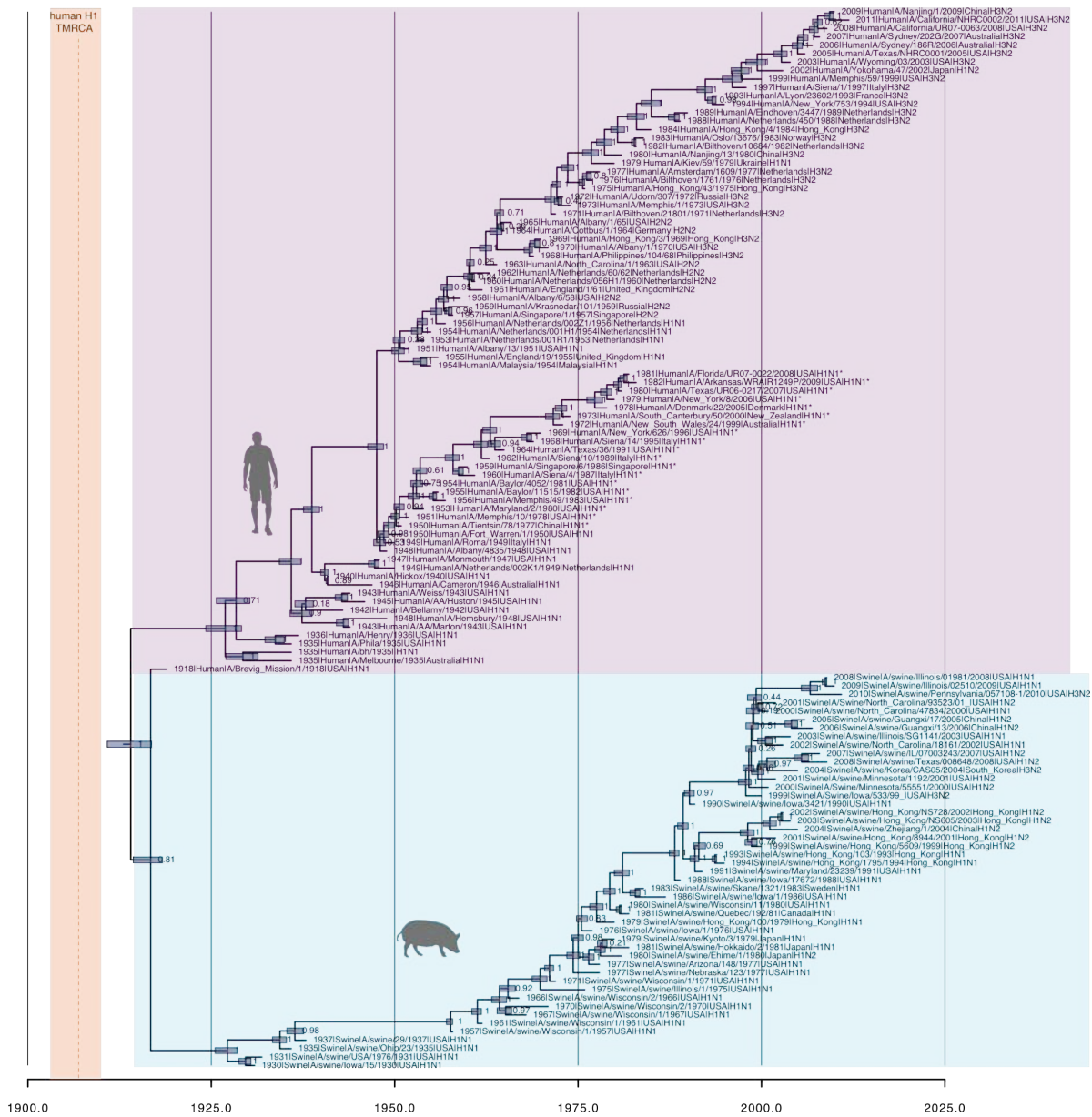
B *PBI*



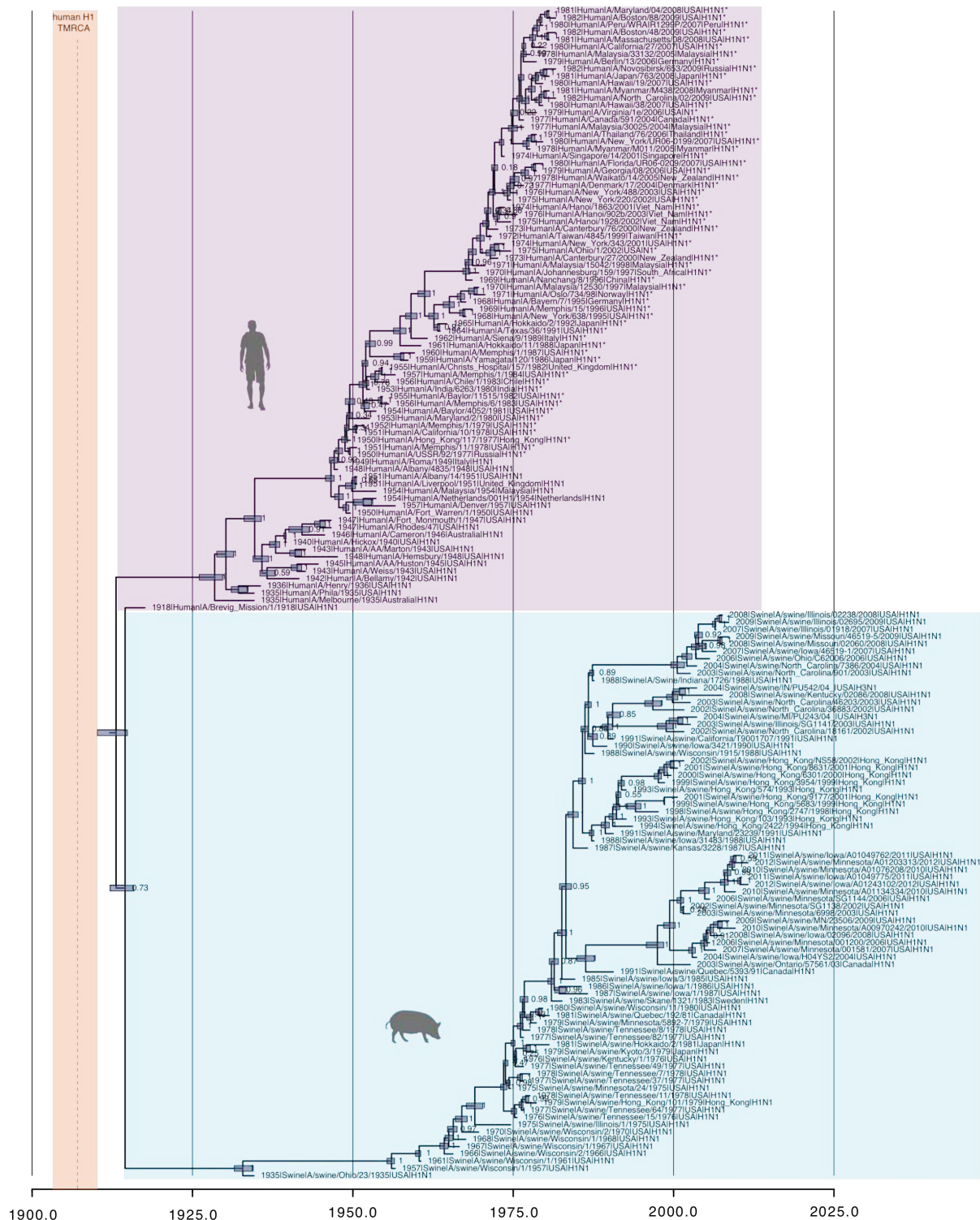


D_{HA} (H1)

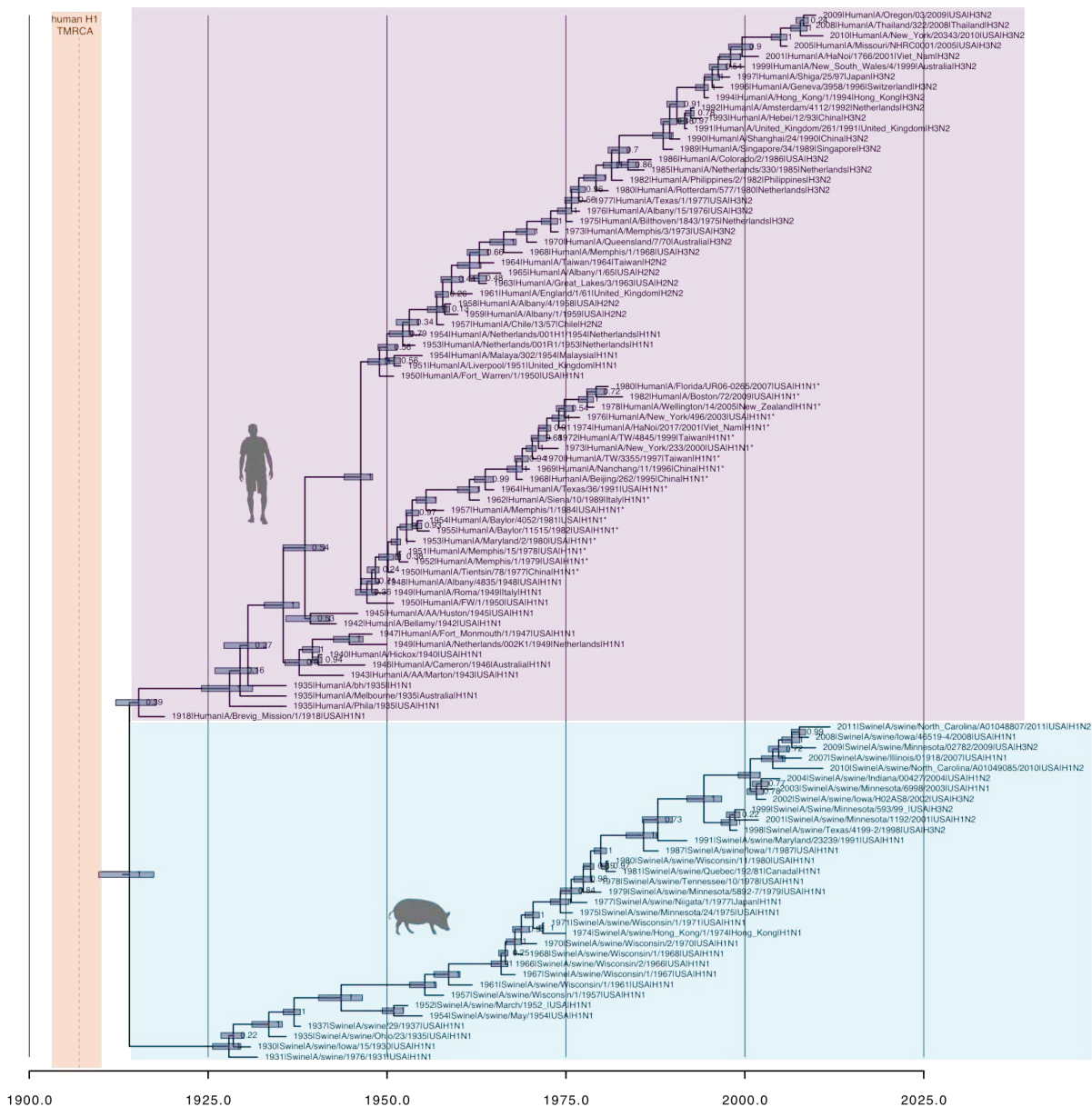




F_{N4}(N1)



G_{M1/2}



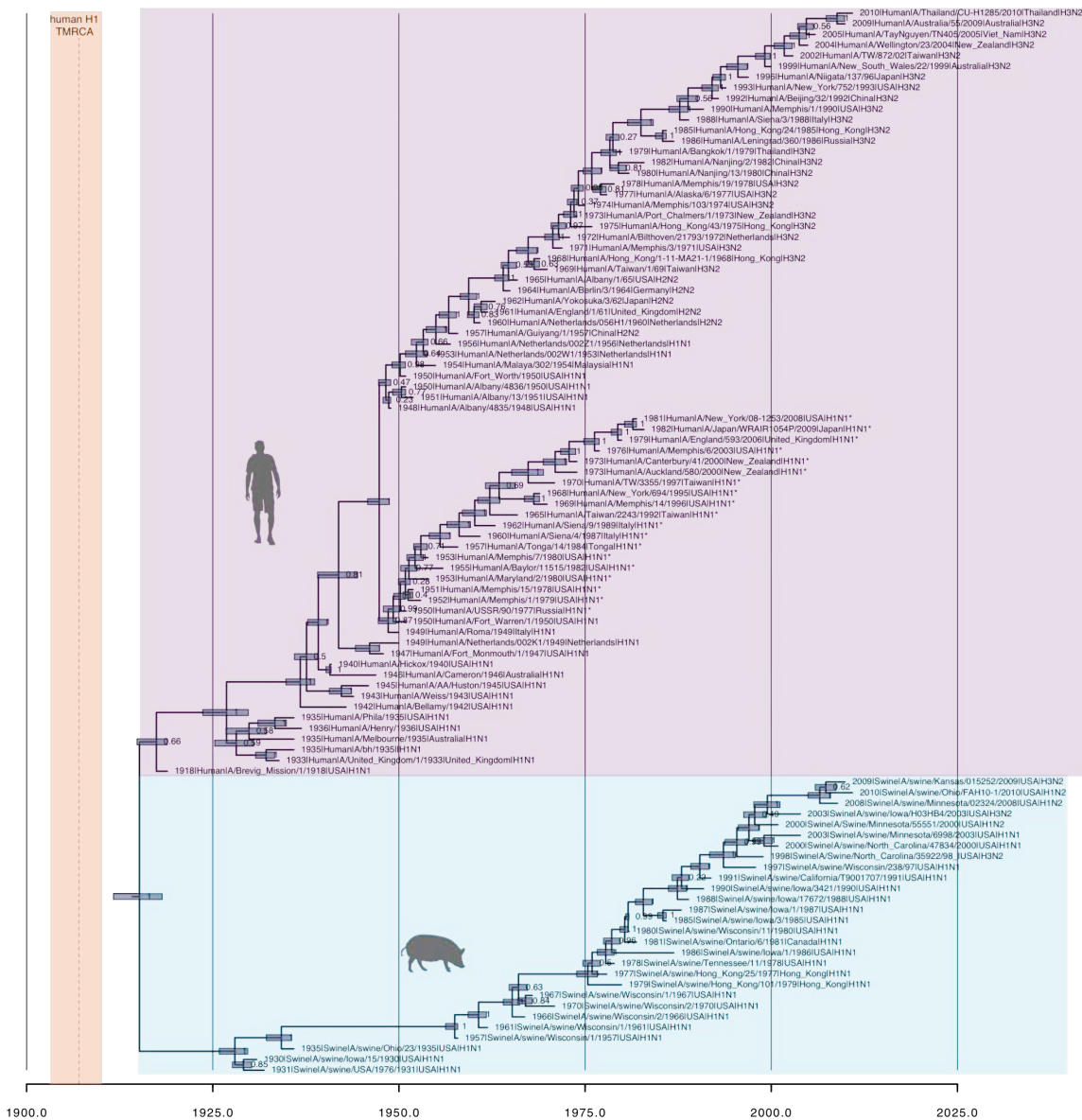


Fig. S12. Schematic diagrams of various host-jumping hypotheses leading to observed H1 HA variation.

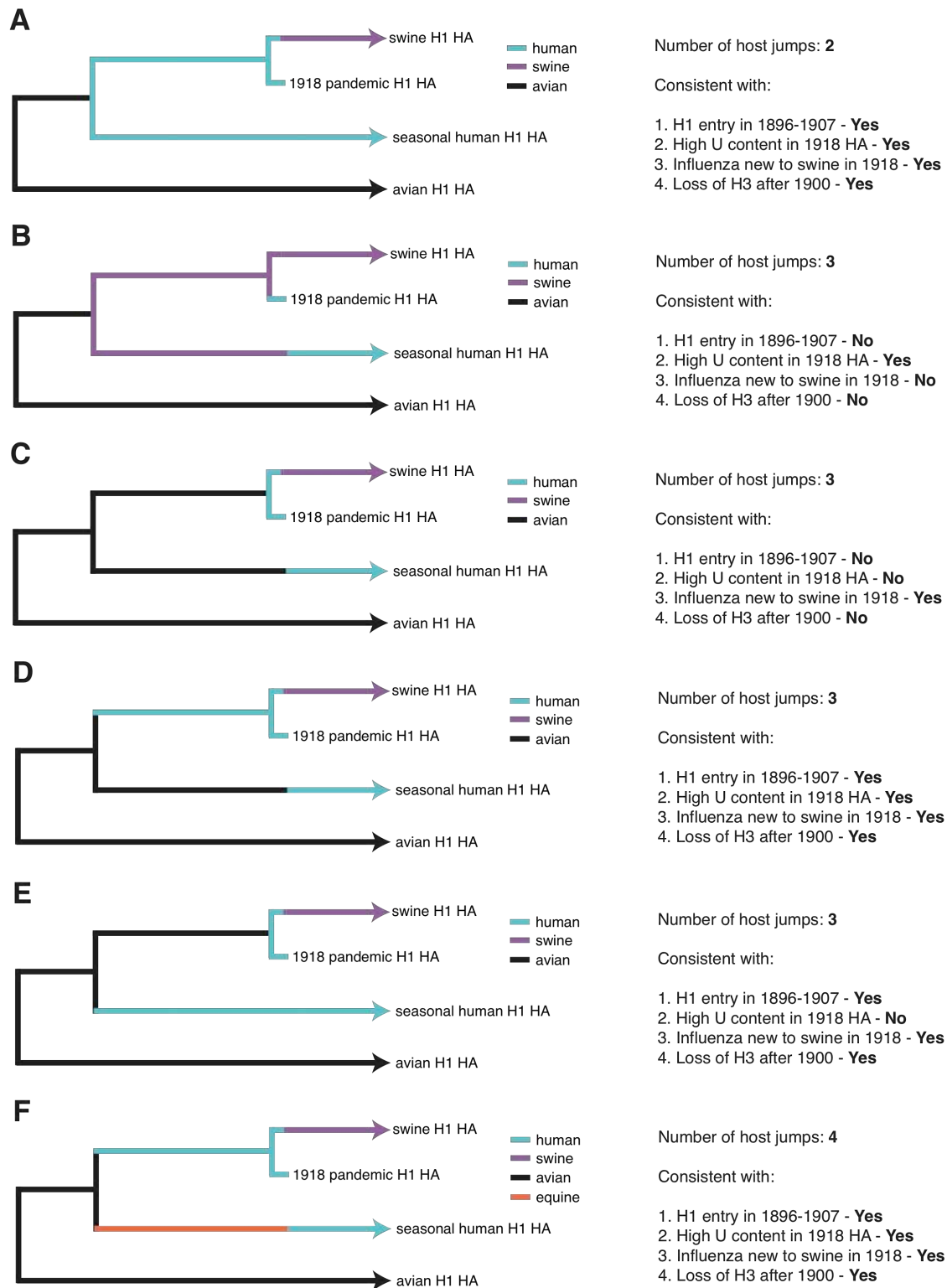


Fig. S13. Maximum clade credibility (MCC) tree of the H1, H2, and H5 subtypes of HA. To the right of the MCC tree are the clade-specific rate distributions (in substitutions/site/year) inferred under the local clock model. The pandemics of 1830-33, 1847-51, and 1889-1893 are shown with gray bars. The posterior probability of each node and 95% CIs on node times are shown. The orange stars indicate points at which H1 or H1-like viruses are assumed to have emerged in humans (in 1830, ~1900, and 2009; the 1977 re-emergence is not marked). The cross represents the putative extinction of the 1830-1889 H1-like HA. That lineage, and the classical swine influenza lineage, represented by dashed branches, are superimposed on the tree for purposes of illustration and were not part of the phylogenetic analysis.

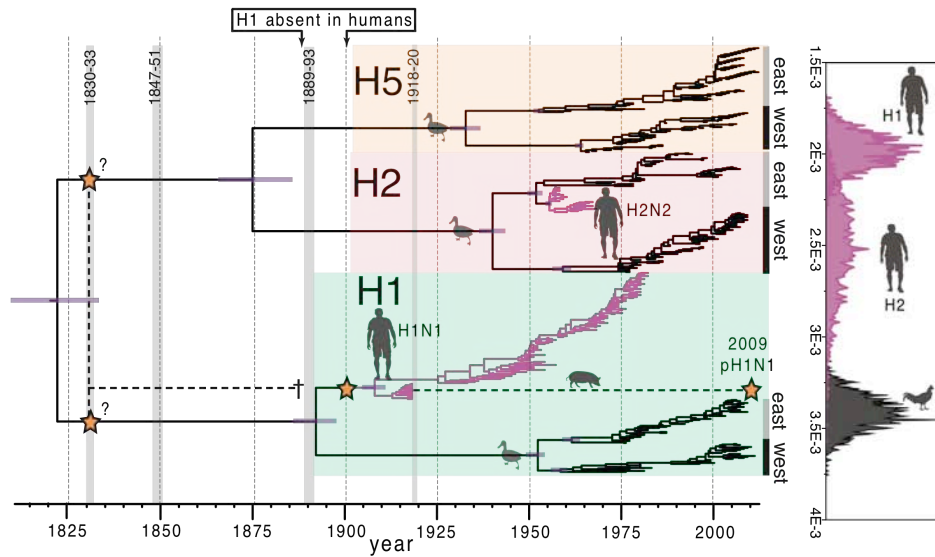
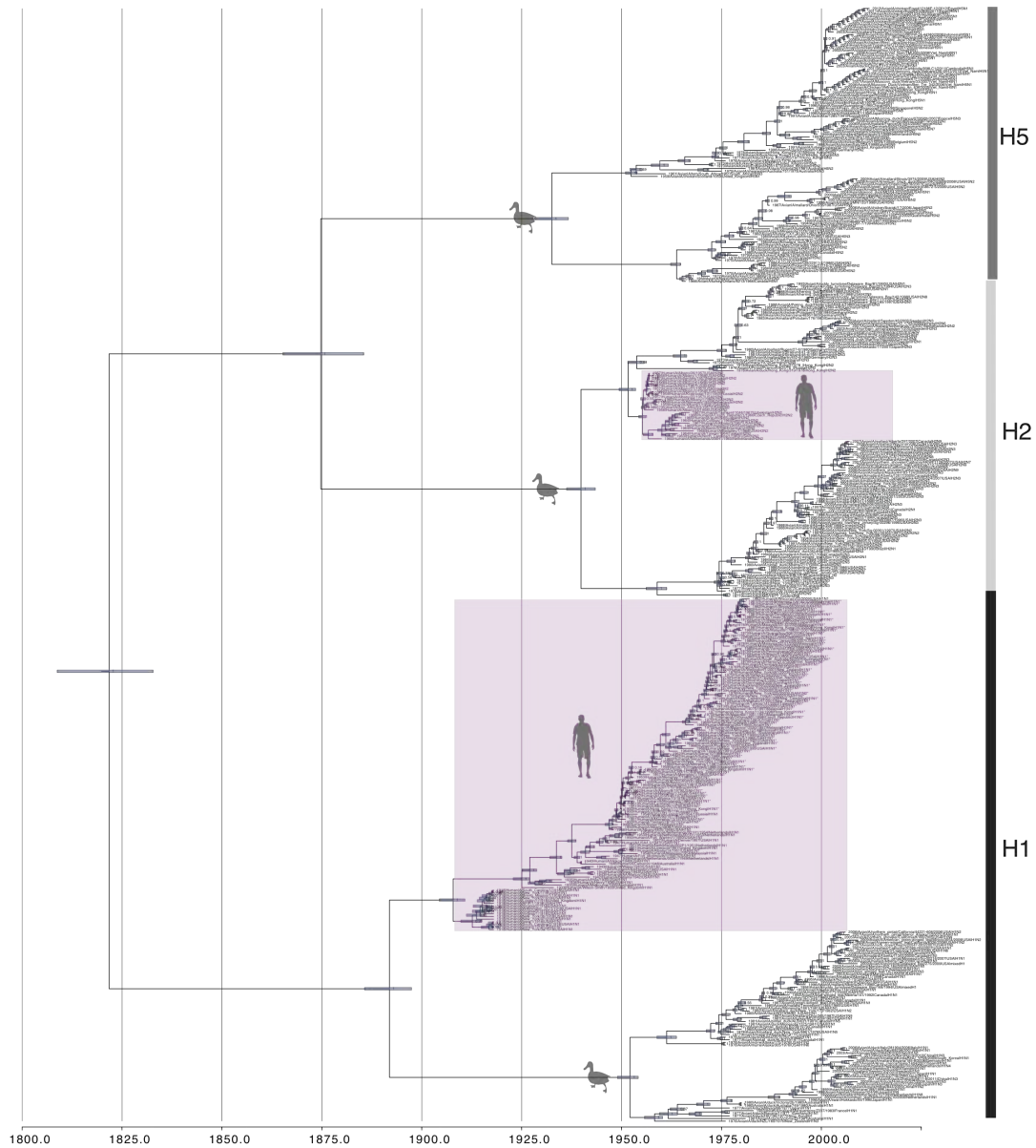


Fig. S14. H1, H2, H5 MCC tree. Branch lengths are in years. Posterior probabilities of each node and 95% CIs on node dates are shown. Human H1 sequences sampled after 1977 (tip-date corrected by -27 years) are marked with an asterisk. **(A)** shared rate for avian H1, H2, and H5. **(B)** independent rates for avian H1, H2, and H5.

A



B

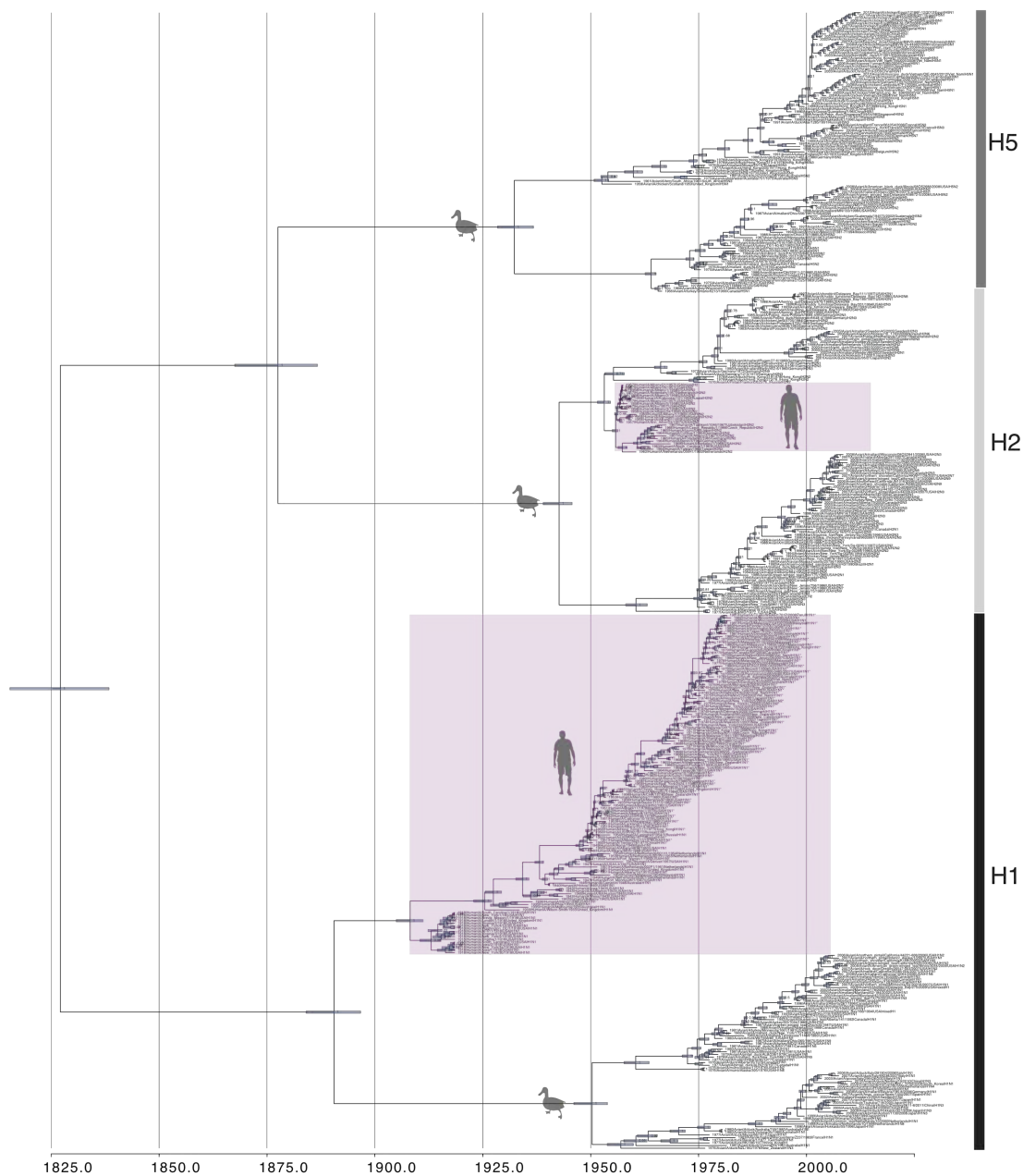


Fig. S15. Fatalities by birth year for H5N1 and H7N9. **(A)** Percentage of fatalities accounted for by 10-year-wide birth year cohorts, for H5N1 and H7N9. Birth year is plotted on the x-axis and the bins are centered on the midpoint of each cohort. Almost all patients with birth years prior to 1968 are expected to have been initially exposed, as children, to a group 1 HA (either H1N1 or H2N2) (blue shading). Most of those born in 1968 and later were exposed first to a group 2 HA (H3N2) (orange shading). **(B)** The same data, but corrected for the percentage of the total population contained in each 10-year birth year cohort in China in 2013 or Indonesia in 2006. See Supplementary text for additional details.

